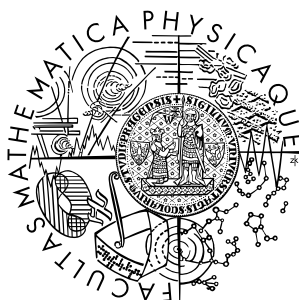


Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE



Iva Zymáková

Ověřování pravděpodobnostních předpovědí

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: Mgr. Zdeněk Hlávka, Ph.D.

Studijní program: obecná matematika

2007

Děkuji Mgr. Zdeňkovi Hlávkovi, Ph.D. za vedení, konzultace a pomoc při vytváření této bakalářské práce a Mgr. Martinovi Betincovi za poskytnutí dat.

Prohlašuji, že jsem svou bakalářskou práci napsala samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 6. září 2007

Iva Zymáková

Obsah

1	Úvod	5
1.1	Historie	5
1.2	Předpovědi, pozorování a vzorek ověřovacích dat	6
2	Faktorizace sdruženého rozdělení předpovědí a pozorování	7
2.1	CR faktorizace	7
2.2	LBR faktorizace	8
2.3	Vztah mezi faktorizacemi	8
3	Veličiny popisující sdružené rozdělení předpovědí a pozorování	9
3.1	Lineární regrese	10
3.2	Míry kvality předpovědí	10
3.3	Rozklady MSE	11
3.4	Rozklad SS	13
3.5	Charakteristiky kvality předpovědí	13
4	Srovnání s SDT přístupem	15
4.1	Klasifikátor, skórová funkce	15
4.2	ROC křivka	15
4.3	Porovnání ROC křivky a LBR faktorizace	16
5	Binormální model a MSE	18
5.1	Binormální model	18
5.2	MSE v binormálním modelu	19
6	Aplikace na konkrétní předpovědi	22
6.1	Sdružené rozdělení	22
6.2	CR faktorizace	23
6.3	LBR faktorizace	23
6.4	Míry kvality předpovědí	25
6.5	Rozklady MSE	25
6.6	Rozklad SS	27
6.7	ROC křivka	27
6.8	Shrnutí	27
	Literatura	29

Název práce: Ověřování pravděpodobnostních předpovědí
Autor: Iva Zymáková
Katedra (ústav): Katedra pravděpodobnosti a matematické statistiky
Vedoucí bakalářské práce: Mgr. Zdeněk Hlávka, Ph.D.
e-mail vedoucího: Zdenek.Hlavka@mff.cuni.cz

Abstrakt: V práci je popsán diagnostický přístup k ověřování pravděpodobnostních předpovědí. Základním prvkem tohoto přístupu je sdružené rozdělení předpovědí a pozorování spolu s příslušnými podmíněnými a marginálními rozděleními. Tato rozdělení a rozklady jejich charakteristik poskytují důležité informace o vztahu mezi předpověďmi a příslušnými pozorováními.

Pro srovnání je v práci zmíněn tzv. SDT přístup (z angl. signal detection theory), jehož hlavním nástrojem pro hodnocení předpovědí jsou ROC (z angl. receiver operator characteristic) křivky. Je nastíněn vztah mezi SDT přístupem a veličinami popisujícími sdružené rozdělení předpovědí a pozorování.

Všechny postupy jsou ilustrovány na předpovědích získaných v roce 2005 na katedře sociologie FF UK v rámci výzkumu „Aktéři, rizika a důvěra“.

Klíčová slova: Ověřování předpovědí, Pravděpodobnostní předpovědi, Diagnostické ověřování předpovědí

Title: Verification of probability forecasts
Author: Iva Zymáková
Department: Department of Probability and Mathematical Statistics
Supervisor: Mgr. Zdeněk Hlávka, Ph.D.
Supervisor's e-mail address: Zdenek.Hlavka@mff.cuni.cz

Abstract: In this work the diagnostic approach to the verification of probability forecasts is described. The basic elements of this approach are the joint distribution of forecasts and observations together with the relevant conditional and marginal distributions. These distributions, their measures and the decompositions of these measures provide useful information about the relationship between the forecasts and the observations.

For comparison the Signal Detection Theory (SDT) approach to the evaluation of forecasts is mentioned in this work, the basic tool in this approach being the Receiver Operator Characteristic (ROC) curve. The relationship between the SDT approach and the measures describing the joint distribution of forecasts and observations is outlined here.

All methods are illustrated by references to the forecasts obtained in „Aktéři, rizika a důvěra“, this research proceeded at The Department of Sociology at the Faculty of Philosophy of The Charles University in Prague in 2005.

Keywords: Forecast verification, Probability forecasts, Diagnostic verification of forecasts

Kapitola 1

Úvod

Diagnostický přístup k ověřování předpovědí je založený na analýze sdruženého rozdělení předpovědí a pozorování (viz [7]). Cílem tohoto přístupu je nalezení silných a slabých stránek předpovědních systémů. Díky tomu může tvůrcům předpovědí poskytnout informace o tom, v čem jsou jimi vytvořené předpovědi dobré a co by se dalo ještě zlepšit.

K nalezení předností a slabých míst předpovědí je dobré zkoumat předpovědi z různých hledisek. Proto má například smysl zkoumat dvě různé faktorizace téhož sdruženého rozdělení (v kapitole 2) nebo různé rozklady střední čtvercové chyby (v kapitole 3).

V kapitole 4 je popsáno hodnocení předpovědí pomocí ROC křivek. Tento (tzv. SDT) přístup je v jistém smyslu podobný LBR faktorizaci sdruženého rozdělení (popsané v kapitole 2). Nevýhodou SDT přístupu je to, že je použitelný pouze v situacích, kdy je množina možných pozorování jen dvouprvková.

ROC křivka se dá spočítat z empirických odhadů podmíněných pravděpodobností $p(f|x = 1)$ a $p(f|x = 0)$ nebo může být odvozena z různých modelů. Jeden takový model (binormální model) je popsán v kapitole 5.

Diagnostický přístup je ilustrován na konkrétních předpovědích v kapitole 6.

1.1 Historie

Prvním, kdo se tématem ověřování předpovědí zabýval, byl meteorolog americké armády seržant John Parker Finley, viz [3]. V roce 1884 uveřejnil v *American Meteorological Journal* článek, ve kterém se pokusil shrnout výsledky předpovědí výskytu tornád. Jako míru úspěšnosti vzal Finley procentuální úspěšnost předpovědí (počet správných předpovědí vydělený počtem všech předpovědí). Došel tak k hodnotě 96.6%.

Tento článek vyvolal diskuzi na téma ověřování předpovědí (tzv. Finleyho aféry). Během 6 měsíců od uveřejnění Finleyho práce byly publikovány články M. H. Doolittle [2], G. K. Gilberta [4] a C. S. Peirce [8], které poukázaly na to, že při stálé předpovědi „nebude tornádo“ by se Finleyho metodou dostala úspěšnost 98.2%. Byly proto navrženy jiné možnosti jak ohodnotit předpovědní systém. Některé z nich byly „znovu objeveny“ o sto let později a mnohé se používají dodnes.

Finleyho aférou a jednotlivými pracemi se podrobněji zabývá [5].

1.2 Předpovědi, pozorování a vzorek ověřovacích dat

Systémy, pomocí nichž jsou vytvářeny předpovědi, budu souhrnně nazývat *předpovědními systémy*. Mohou to být různé predikční modely a metody, ale třeba i prognostikové nebo věštcí. Ověřováním předpovědi rozumíme hodnocení a porovnávání jednotlivých předpovědních systémů. Předpovědní systémy se obvykle značí velkými písmeny F, G, \dots . Konkrétní předpovědi vytvořené předpovědními systémy F, G, \dots se pak značí malými písmeny f, g, \dots .

Aby se předpovědní systémy daly ohodnotit nebo porovnat, musíme mít k dispozici skutečná pozorování, příslušná k jednotlivým předpovědím. Množiny možných pozorování, příslušné k daným předpovědím, se značí písmeny X, Y, \dots . Jednotlivé hodnoty pozorování jsou x, y, \dots .

Předpovědi se dají vyjádřit buď v pravděpodobnostním nebo v nepravděpodobnostním tvaru. Pravděpodobnostní předpovědi se snaží určit pravděpodobnost, s jakou určitý jev nastane (nebo nenastane). Nepravděpodobnostní předpovědi pouze říkají, jestli událost nastane nebo ne. V případě nepravděpodobnostních předpovědí jsou množiny možných předpovědí a pozorování totožné. Celá tato práce je zaměřena na předpovědi pravděpodobnostní. O vztahu mezi pravděpodobnostními a nepravděpodobnostními předpověďmi se zmiňuji v kapitole 4.

Vzorek ověřovacích dat, který přísluší předpovědnímu systému F a předpovídaným situacím X , je systém n dvojic předpovědi a pozorování $\{(f^{(i)}, x^{(i)}), i = 1, \dots, n\}$, kde n je velikost vzorku. Dvojice $(f^{(i)}, x^{(i)})$ a $(f^{(j)}, x^{(j)})$, pro $i \neq j$, nemusejí být nutně rozdílné.

Někdy se zavádí další značení. Označme n^f (resp. n^x) počet různých hodnot předpovědi (resp. pozorování), p_{ij} bude sdružená relativní četnost hodnoty předpovědi f_i spolu s pozorováním x_j ve vzorku ověřovacích dat (pro $i = 1, \dots, n^f$ a $j = 1, \dots, n^x$). Matice $P = (p_{ij})_{i=1, j=1}^{n^f, n^x}$ obsahuje sdružené relativní četnosti všech (f, x) -párů. Tyto četnosti se obvykle berou jako sdružené pravděpodobnosti předpovědi f_i spolu s pozorováním x_j . Tedy $p_{ij} = P(F = f_i, X = x_j)$. Pravděpodobnosti p_{ij} se dají chápat jako parametry sdruženého rozdělení předpovědi a pozorování, které budu značit $p(f, x)$.

V dalším budu uvažovat pouze situace, kde je množina možných pozorování jen dvouprvková, tj. $n^x = 2$. (Buď předpovídaná událost nastane nebo nenastane.)

Poznámka: Ve skutečnosti hodnoty $f^{(1)}, \dots, f^{(n)}$ obvykle vznikají „zaokrouhlením“ nějakých čísel f^1, \dots, f^n na hodnoty z množiny $\{f_1, \dots, f_{n^f}\}$. Původní hodnoty f^1, \dots, f^n jsou náhodným výběrem ze spojitého rozdělení náhodné veličiny F na intervalu $[0, 1]$. Hodnoty x^1, \dots, x^n tvoří výběr z alternativního rozdělení. Prakticky se s nimi ale zachází jako s náhodnými veličinami s diskrétním rozdělením, které je určeno relativními četnostmi jednotlivých hodnot f_1, \dots, f_{n^f} (resp. x_1, \dots, x_{n^x}).

Kapitola 2

Faktorizace sdruženého rozdělení předpovědí a pozorování

Symbol $p(f, x)$ bude označovat sdružené rozdělení předpovědí (získaných pomocí předpovědního systému F) a pozorování. Předpovědi i pozorování se obvykle značí malými písmeny, bez ohledu na to, představují-li náhodné veličiny nebo nějaké konkrétní hodnoty.

f_i budou jednotlivé pravděpodobnostní předpovědi (tedy $f_i \in [0, 1]$) a x_i jim příslušná pozorování. Vzhledem k tomu, že uvažují pouze situace, kde je množina možných pozorování jen dvouprvková, je $x_i \in \{0, 1\}$. V případě, že je $x_i = 1$, předpovídaná událost nastala a v případě, že $x_i = 0$, nenastala.

Informace uložené ve sdruženém rozdělení $p(f, x)$ se zpřístupní, pokud $p(f, x)$ rozložíme na podmíněné a marginální rozdělení. Dvě možné faktorizace jsou:

1. CR faktorizace (calibration-refinement factorization) (2.1) a
2. LBR faktorizace (likelihood-base rate factorization) (2.2).

2.1 CR faktorizace

CR faktorizace je vyjádřena vztahem

$$p(f, x) = p(x|f) \cdot p(f). \quad (2.1)$$

$p(x|f)$ (první člen na pravé straně rovnosti (2.1)) je podmíněné rozdělení výskytu pozorované události při jednotlivých předpovědích. Vypovídá o tzv. *kalibraci* (calibration). Vzorek je perfektně kalibrován, pokud je pro každou hodnotu pravděpodobnostní předpovědi relativní četnost předpovídaného jevu $p(x = 1|f)$ rovna hodnotě této předpovědi ($\forall f p(x = 1|f) = f$). Graf podmíněné střední hodnoty $\mu_{x|f} = p(x = 1|f)$ jako funkce f se nazývá kalibrační diagram. Perfektně kalibrován vzorek dat má všechny hodnoty na diagonále $\mu_{x|f} = f$, čím více se hodnoty $\mu_{x|f}$ vzdalují od této úsečky, tím je vzorek hůř kalibrován.

Marginální rozdělení předpovědí $p(f)$ je tzv. *ostrost* (sharpness, refinement). Vzorek pravděpodobnostních předpovědí je úplně ostrý, pokud pravděpodobnostní předpovědi nabývají jen hodnot 0 a 1.

Poznámka: Úplně ostré předpovědi by zřejmě byly špatně kalibrované (vzhledem k nemožnosti předpovědět budoucnost s jistotou). Ze stejného důvodu vykazují dobře kalibrované předpovědi jen mírný stupeň ostrosti.

2.2 LBR faktorizace

LBR faktorizace je vyjádřena vztahem

$$p(f, x) = p(f|x) \cdot p(x). \quad (2.2)$$

Podmíněné pravděpodobnosti, které určují $p(f|x)$, se nazývají *věrohodnosti* (likelihoods). Věrohodnosti ukazují, jakou měrou daná předpověď upřednostňuje různé hodnoty pozorování. Když je pro určitou předpověď jeden z výrazů $p(f|x = 1)$ a $p(f|x = 0)$ nulový, pak řekneme, že tato předpověď je úplně diskriminující. Čím víc se poměr věrohodností $p(f|x = 1)/p(f|x = 0)$ liší od jedné, tím je předpověď f víc diskriminující. Obrázek s grafy věrohodnostních funkcí $p(f|x = 0)$ a $p(f|x = 1)$ (tzv. likelihood diagram) dává vizuální představu o stupni diskriminace.

Marginální distribuční funkce $p(x)$ je jediná komponenta obou faktorizací, která nezávisí na předpovědích. Je to spíš charakteristika předpovídané situace než předpovědního systému. Hodnoty $p(x)$ (tj. $p(x = 1)$ a $p(x = 0)$) se anglicky nazývají *base rates*, což by se do češtiny dalo přeložit jako základní poměry.

Poznámka: Konstantní předpovědi rovnající se base rate ($p(x = 1)$) jsou obvykle velice dobře kalibrované, ale naprosto neostré.

2.3 Vztah mezi faktorizacemi

Čtyři komponenty faktorizací (2.1) a (2.2) sdruženého rozdělení zahrnují čtyři různé charakteristiky předpovědního systému. Přesto je mezi nimi určitá závislost. Zřejmě platí, že

$$p(x|f) \cdot p(f) = p(f|x) \cdot p(x). \quad (2.3)$$

Například, je-li předpovědní systém perfektně kalibrovaný ($\forall f p(x = 1|f) = f$) a úplně ostrý ($\forall f f = 0$ nebo $f = 1$), pak musí být věrohodnosti úplně diskriminující. Naopak to platit nemusí (viz [6]).

K popsání sdruženého rozdělení f a x můžeme použít vhodné charakteristiky jeho faktorizací. Je výhodné shrnout rysy rozdělení do několika konkrétních veličin.

Kapitola 3

Veličiny popisující sdružené rozdělení předpovědí a pozorování

Předpovědi i pozorování se chápou jako náhodné veličiny, které mají diskrétní rozdělení (viz kapitola 1.2).

Důležitou úlohu při popisování sdruženého rozdělení předpovědí a pozorování hrají obvyklé charakteristiky, jakými jsou střední hodnoty podmíněných i marginálních rozdělení:

$$\mu_x = \sum_x xp(x), \quad (3.1)$$

$$\mu_f = \sum_f fp(f), \quad (3.2)$$

$$\mu_{x|f} = \sum_x xp(x|f), \quad (3.3)$$

$$\mu_{f|x} = \sum_f fp(f|x), \quad (3.4)$$

rozptyly marginálních rozdělení:

$$\sigma_x^2 = E_x(x - \mu_x)^2, \quad (3.5)$$

$$\sigma_f^2 = E_f(f - \mu_f)^2 \quad (3.6)$$

a korelační koeficient mezi předpověďmi a pozorováními:

$$\rho_{fx} = \frac{\text{cov}(f, x)}{\sqrt{\sigma_f^2 \sigma_x^2}}. \quad (3.7)$$

Vzhledem k tomu, že uvažují pouze situace, kde je množina možných pozorování jen dvouprvková, je střední hodnota pozorování, ve vzorci (3.1), rovna

$$\mu_x = P(x = 1). \quad (3.8)$$

3.1 Lineární regrese

K popsání některých vlastností podmíněných rozdělení $p(x|f)$ a $p(f|x)$ je dobré použít následující lineární regresní modely. Hledáme-li pro $\mu_{x|f}$ nejlepší lineární aproximaci pomocí předpovědi f , dostáváme

$$\hat{\mu}_{x|f} = a + bf, \quad (3.9)$$

kde koeficienty a a b jsou vyjádřeny pomocí středních hodnot, rozptylů a korelačního koeficientu, tj. $a = \mu_x - b\mu_f$ a $b = (\sigma_x/\sigma_f)\rho_{fx}$. V souvislosti s (3.9) jsou předpovědi perfektně kalibrované právě tehdy, když $a = 0$ a $b = 1$.

Analogicky pro $\mu_{f|x}$ máme

$$\hat{\mu}_{f|x} = c + dx, \quad (3.10)$$

kde $c = \mu_f - d\mu_x$ a $d = (\sigma_f/\sigma_x)\rho_{fx}$.

3.2 Míry kvality předpovědí

K dalšímu objasnění vztahu mezi předpověďmi a pozorováními je užitečné spočítat některé míry kvality předpovědí (v angličtině označované jako performance measures) – střední chybu (ME), střední čtvercovou chybu (MSE) a skill score (SS).

Střední chyba

(ME, mean error) je mírou *nepodmíněného vychýlení* (unconditional bias) a je definována vztahem

$$\text{ME}(f, x) = \mu_f - \mu_x. \quad (3.11)$$

Střední čtvercová chyba

(MSE, mean square error, někdy bývá nazývána Brierovo skóre) je mírou *celkové správnosti* (accuracy). MSE je definována vztahem

$$\text{MSE}(f, x) = \sum_f \sum_x (f - x)^2 p(f, x). \quad (3.12)$$

Skill score

(SS) je míra celkové správnosti předpovědí, které nás zajímají, v poměru k celkové správnosti předpovědí založených na nějakém srovnávacím standardu. Například SS založený na MSE a na střední hodnotě rozdělení pozorování je definován vztahem

$$\text{SS}(f, x) = 1 - [\text{MSE}(f, x) / \text{MSE}(\mu_x, x)]. \quad (3.13)$$

$\text{MSE}(\mu_x, x)$ ve vzorci (3.13) je chápána jako MSE konstantní předpovědi o velikosti střední hodnoty pozorování μ_x a je rovna rozptylu pozorování σ_x^2 .

3.3 Rozklady MSE

MSE se dá rozložit několika způsoby.

Základní rozklad:

$$\text{MSE}(f, x) = (\mu_f - \mu_x)^2 + \sigma_f^2 + \sigma_x^2 - 2\sigma_f\sigma_x\rho_{fx} \quad (3.14)$$

Důkaz:

$$\begin{aligned} & (\mu_f - \mu_x)^2 + \sigma_f^2 + \sigma_x^2 - 2\sigma_f\sigma_x\rho_{fx} \\ &= (\mu_f - \mu_x)^2 + \sigma_f^2 + \sigma_x^2 - 2\text{cov}(f, x) \\ &= (\mathbf{E} f)^2 - 2(\mathbf{E} f)(\mathbf{E} x) + (\mathbf{E} x)^2 + \mathbf{E} f^2 - (\mathbf{E} f)^2 + \mathbf{E} x^2 - (\mathbf{E} x)^2 \\ &\quad - 2\mathbf{E}(fx) + 2(\mathbf{E} f)(\mathbf{E} x) \\ &= \mathbf{E} f^2 - 2\mathbf{E}(fx) + \mathbf{E} x^2 \\ &= \mathbf{E}(f - x)^2 \\ &= \sum_f \sum_x (f - x)^2 p(f, x) = \text{MSE}(f, x) \end{aligned}$$

□

První člen na pravé straně rovnosti (3.14) je druhá mocnina střední chyby, neboli míra nepodmíněného vychýlení. Výraz $\sigma_f^2 + \sigma_x^2 - 2\sigma_f\sigma_x\rho_{fx}$ se dá chápat jako rozptyl chyb předpovědi ($\text{var}(f - x)$).

CR rozklad:

$$\text{MSE}(f, x) = \sigma_x^2 + \mathbf{E}_f (\mu_{x|f} - f)^2 - \mathbf{E}_f (\mu_{x|f} - \mu_x)^2 \quad (3.15)$$

Důkaz: Pro podmíněnou střední hodnotu $\mu_{x|f}$ platí

$$\mu_{x|f} = \sum_x x p(x|f) = \sum_x x \frac{p(f, x)}{p(f)}.$$

To dosadíme do pravé strany rovnosti (3.15):

$$\begin{aligned} & \sigma_x^2 + \mathbf{E}_f (\mu_{x|f} - f)^2 - \mathbf{E}_f (\mu_{x|f} - \mu_x)^2 \\ &= \mathbf{E} x^2 - (\mathbf{E} x)^2 + \mathbf{E}_f (\mu_{x|f}^2 - 2\mu_{x|f}f + f^2 - \mu_{x|f}^2 + 2\mu_{x|f}\mu_x - \mu_x^2) \\ &= \mathbf{E} x^2 - (\mathbf{E} x)^2 + \mathbf{E}_f (f^2 - 2\mu_{x|f}f + 2\mu_{x|f}\mu_x - \mu_x^2) \\ &= \mathbf{E} x^2 - (\mathbf{E} x)^2 + \mathbf{E} f^2 - 2\mathbf{E}_f \left(\sum_x x \frac{p(f, x)}{p(f)} f \right) \\ &\quad + 2\mathbf{E}_f \left(\sum_x x \frac{p(f, x)}{p(f)} \mathbf{E} x \right) - (\mathbf{E} x)^2 \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} x^2 - 2(\mathbb{E} x)^2 + \mathbb{E} f^2 - 2 \sum_f \sum_x \left(f x \frac{p(f, x)}{p(f)} p(f) \right) \\
&\quad + 2 \sum_f \sum_x \left(x \mathbb{E} x \frac{p(f, x)}{p(f)} p(f) \right) \\
&= \mathbb{E} x^2 - 2(\mathbb{E} x)^2 + \mathbb{E} f^2 - 2 \mathbb{E} (fx) + 2(\mathbb{E} x)^2 \\
&= \mathbb{E} (f^2 - 2fx + x^2) = \mathbb{E} (f - x)^2 \\
&= \sum_f \sum_x (f - x)^2 p(f, x) = \text{MSE}(f, x)
\end{aligned}$$

□

Výraz σ_x^2 v rovnosti (3.15) je rozptyl rozdělení pozorování, nezáleží na předpovědích, dá se chápat jako MSE konstantní předpovědi o velikosti střední hodnoty pozorování μ_x .

Veličina $\mathbb{E}_f(\mu_{x|f} - f)^2$ měří kalibraci. Tento člen je nulový jen pro perfektně kalibrované předpovědi. Jinak přispívá ke zvětšení MSE.

Třetí člen na pravé straně rovnosti (3.15) je mírou *rezoluce*, *rozlišení* (resolution). Obvykle je žádoucí, aby pro „velké“ předpovědi byla podmíněná relativní četnost výskytu určité události větší a pro „malé“ předpovědi menší než nepodmíněná relativní četnost jejího výskytu.

$\mathbb{E}_f(\mu_{x|f} - \mu_x)^2$ dává kvantitativní míru rezoluce. Zmizí pro naprosto nerozlišující předpovědi (tj. když $\mu_x = \mu_{x|f}$ pro všechna f). Jinak přispívá ke snížení MSE.

Pro perfektně kalibrované předpovědi platí $\mu_{x|f} = f$ a $\mu_x = \mu_f$. Takže rezoluční člen je roven rozptylu předpovědí ($\mathbb{E}_f(\mu_{x|f} - \mu_x)^2 = \mathbb{E}_f(f - \mu_f)^2 = \sigma_f^2$). Rozptyl předpovědí je ale mírou ostrosti. Rezoluce a ostrost jsou pak tedy (pro perfektně kalibrované předpovědi) shodné.

Tedy z CR rozkladu (3.15) vidíme, že se střední čtvercová chyba rovná rozptylu pozorování zvětšenému o „pokutu“ přímo závislou na stupni nekalibrovanosti a zmenšenému o „vylepšení“, které závisí na stupni rezoluce (rozlišení).

Řekneme, že předpovědi vykazují kladnou *úspěšnost* (skill), pokud vylepšení založené na rezoluci překoná zhoršení z nekalibrovanosti.

LBR rozklad:

$$\text{MSE}(f, x) = \sigma_f^2 + \mathbb{E}_x(\mu_{f|x} - x)^2 - \mathbb{E}_x(\mu_{f|x} - \mu_f)^2 \quad (3.16)$$

První člen na pravé straně rovnosti (3.16) je rozptyl předpovědí.

Poznámka: Mohlo by se zdát, že je žádoucí zmenšení variability předpovědí (tj. že se zmenšujícím se σ_f^2 bude klesat i MSE). Ale není tomu tak. Pokud zmenšíme σ_f^2 , pak se $\mathbb{E}_x(\mu_{f|x} - x)^2$ zvětší a $\mathbb{E}_x(\mu_{f|x} - \mu_f)^2$ zmenší, a ke zlepšení nedojde.

Druhý člen na pravé straně rovnosti (3.16) je mírou tzv. *nepodmíněného vychýlení*. Je nulový, jen když podmíněná střední hodnota předpovědí za podmínky, že předpovídaná událost nastala, je rovna 1, jinak je rovna 0. Jinak $\mathbb{E}_x(\mu_{f|x} - x)^2$ zvětšuje MSE.

Poslední člen $E_x(\mu_{f|x} - \mu_f)^2$ je mírou *diskriminace*. Je nulový, když podmíněná střední hodnota předpovědi za podmínky, že předpovídaná událost nastala, je rovna celkové střední hodnotě předpovědi. Jinak přispívá ke zmenšení MSE.

Z LBR rozkladu tedy plyne, že je vhodné, aby se střední hodnota předpovědi pro $x = 0$ a $x = 1$ co nejvíc lišila od base rate ($p(x = 0)$ a $p(x = 1)$) a aby se blížila ke krajním bodům intervalu $[0, 1]$.

3.4 Rozklad SS

Skill score se dá rozložit pomocí základního rozkladu MSE:

$$SS(f, x) = \rho_{fx}^2 - [\rho_{fx} - (\sigma_f/\sigma_x)]^2 - [(\mu_f - \mu_x)/\sigma_x]^2 \quad (3.17)$$

Člen ρ_{fx}^2 je druhá mocnina korelačního koeficientu, neboli míra stupně lineární závislosti mezi f a x .

Z lineárního regresního modelu (3.9) plyne, že výraz $[\rho_{fx} - (\sigma_f/\sigma_x)]^2$ je mírou nekalibrovanosti předpovědi. Tento člen je nulový jen pro perfektně kalibrované předpovědi. Výraz $[(\mu_f - \mu_x)/\sigma_x]^2$ měří nepodmíněné vychýlení předpovědi. Je nulový jen pro nevychýlené předpovědi. Jinak druhý a třetí člen v rozkladu SS (3.17) přispívají ke zmenšení skill score.

3.5 Charakteristiky kvality předpovědí

Charakteristiky kvality předpovědí spolu s veličinami, které je měří, jsou shrnuté v tabulce 3.1.

Tyto charakteristiky jsou spočtené a znázorněné pro konkrétní předpovědi v kapitole 6.

Tabulka 3.1: Charakteristiky kvality předpovědí (s grafy a veličinami, které příslušnou charakteristiku měří) – převzato z [6] a [7].

Název	Popis	Grafy a míry
Kalibrace (calibration, reliability)	míra podobnosti mezi podmíněnou relativní četností pozorování $\mu_{x f}$ a danou hodnotou předpovědi f (průměr přes všechny hodnoty předpovědí)	kalibrační diagram; $E_f(\mu_{x f} - f)^2$ – míra nekalibrovanosti; $[\rho_{fx} - (\sigma_f/\sigma_x)]^2$ – míra nekalibrovanosti
Ostrost (refinement, sharpness)	charakteristika popisující, jak jsou pravděpodobnostní předpovědi blízko u 0 a u 1	graf $p(f)$; σ_f^2 – rozptyl předpovědí
Rezoluce, rozlišení (resolution)	míra, s jakou se podmíněné relativní četnosti $\mu_{x f}$ liší od base rate μ_x (průměr přes všechny hodnoty předpovědí)	attributes diagram; $E_f(\mu_{x f} - \mu_x)^2$ – míra rezoluce
Sdružení (association)	síla lineární závislosti mezi předpověďmi a pozorováními	ρ_{fx} – míra asociace
Diskriminace (discrimination)	charakteristika popisující, jak se $\mu_{f x}$ liší od μ_f , to jak předpovědi upřednostňují mezi případy, kdy $x = 0$ a kdy $x = 1$	likelihood diagram – graf $p(f x=1)$ a $p(f x=0)$; $p(f x=1)/p(f x=0)$ – poměry věrohodností; $E_x(\mu_{f x} - \mu_f)^2$ – míra diskriminace;
Nejistota (uncertainty)	charakteristika popisující variabilitu pozorování	σ_x^2 – rozptyl pozorování
Podmíněné vychýlení (conditional bias)	stupeň podobnosti mezi $\mu_{f x}$ a x (průměr přes všechny možné hodnoty pozorování)	$E_x(\mu_{f x} - x)^2$ – míra podmíněného vychýlení
Nepodmíněné vychýlení (unconditional bias)	rozíl mezi střední hodnotou předpovědi μ_f a střední hodnotou pozorování μ_x (base rate)	$ME(f, x) = \mu_f - \mu_x$; $[(\mu_f - \mu_x)/\sigma_x]^2$ – míra nepodmíněného vychýlení
Celková správnost (accuracy)	celková míra podobnosti mezi předpověďmi a pozorováními	MSE (Brier score)
Úspěšnost (skill)	poměr celkové správnosti předpovědi vzhledem ke správnosti předpovědi založených na nějakém srovnávacím standardu (např. konstantních předpovědí s hodnotou base rate)	SS (Skill score); ρ_{fx}^2 – míra úspěšnosti perfektně kalibrovaných předpovědí

Kapitola 4

Srovnání s SDT přístupem

Způsob ověřování předpovědí založený na popisu a analýze sdruženého rozdělení předpovědí a pozorování, který je popsán v kapitolách 2 a 3, se nazývá diagnostický přístup.

Někdy se ale pro hodnocení a porovnávání předpovědí používají metody založené na tzv. SDT přístupu (z anglického signal detection theory). Základním nástrojem SDT přístupu jsou ROC křivky, viz [1].

4.1 Klasifikátor, skórová funkce

Předpokládejme, že se množina sledovaných objektů skládá ze dvou disjunkt-ních tříd \mathcal{X}_1 (kdy předpovídaná událost, tzv. *podmínka*, nastane) a \mathcal{X}_0 (podmínka nenastane). Označíme-li stavový prostor symbolem \mathcal{X} , pak $\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1$, $\mathcal{X}_0 \cap \mathcal{X}_1 = \emptyset$. Jednotlivé sledované objekty pak budou značit $w \in \mathcal{X}$. Přítomnost podmínky indikuje binární náhodná veličina X (skutečná pozorování) svou realizací na každém z objektů. Tedy $X(w) = j$ pro $w \in \mathcal{X}_j$, $j \in \{0, 1\}$.

Reálná funkce $F : \mathcal{X} \rightarrow \mathbf{R}$ se nazývá *skórová funkce*. Pokud je oborem hodnot skórové funkce jen interval $[0, 1]$, pak hodnoty funkce F můžeme chápat jako pravděpodobnostní předpovědi.

Nepravděpodobnostní předpověď (*klasifikátor*) lze chápat jako odhad $G = \hat{X}$ skutečné třídní příslušnosti X . Klasifikátory vyjádřené pomocí skórové funkce F a prahové hodnoty θ rozřazují objekty do tříd podle pravidla:

$$\begin{aligned} F(w) \leq \theta &\Rightarrow G(w) = 0, \\ F(w) > \theta &\Rightarrow G(w) = 1. \end{aligned} \tag{4.1}$$

Klasifikátor G je tak funkcí F (klasifikační metody) a prahové hodnoty θ .

4.2 ROC křivka

Označme $\text{FPR}(\theta) = \text{P}(G(F, \theta) = 1 | X = 0)$ takzvaný *false positive rate* a $\text{TPR}(\theta) = \text{P}(G(F, \theta) = 1 | X = 1)$ *true positive rate*. $\text{FPR}(\theta)$ se dá interpretovat jako pravděpodobnost chyby prvního druhu, $\text{TPR}(\theta)$ jako síla testu.

V ROC grafu se na vodorovnou osu vynáší $FPR(\theta)$ a na svislou $TPR(\theta)$. Množina $[0, 1]^2 = \{(FPR, TPR) | FPR, TPR \in [0, 1]\}$ se nazývá *ROC prostor*.

Klasifikátor zavedený vztahem (4.1) poskytne pro pevné $\theta \in \mathbf{R}$ jednu konkrétní klasifikaci objektů. Ta odpovídá jednomu bodu v ROC prostoru. Známeli $FPR(\theta)$ a $TPR(\theta)$ pro všechna $\theta \in \mathbf{R}$, získáme v ROC prostoru křivku, která se označuje jako *ROC křivka*.

Obecně by se dalo říct, že klasifikátor G je tím lepší, čím je menší vzdálenost jeho ROC křivky od bodu $[0, 1]$ v ROC prostoru. Často se pro hodnocení klasifikátorů používá veličina *AUC* (area under ROC curve), vyjadřující obsah plochy pod grafem ROC křivky. Hodnocení a porovnávání klasifikátorů je popsáno v [1].

4.3 Porovnání ROC křivky a LBR faktorizace

ROC křivka reprezentuje vztah mezi $p(f|x = 1)$ a $p(f|x = 0)$. Je zde tedy určitá podobnost mezi ROC křivkou a LBR faktorizací sdruženého rozdělení $p(f, x)$.

Když je F skórová funkce, která nabývá pouze hodnot z intervalu $[0, 1]$, potom ji můžeme považovat za pravděpodobnostní předpověď.

ROC křivku tvoří body $[FPR(\theta), TPR(\theta)]$ (pro $\forall \theta \in [0, 1]$), které jsou definovány vztahem

$$\begin{aligned} FPR(\theta) &= P(G(F, \theta) = 1 | X = 0) = P(F > \theta | X = 0), \\ TPR(\theta) &= P(G(F, \theta) = 1 | X = 1) = P(F > \theta | X = 1). \end{aligned} \quad (4.2)$$

Každý bod ROC křivky odpovídá jednomu $\theta \in [0, 1]$ a hodnoty $FPR(\theta)$ a $TPR(\theta)$ se ve smyslu LBR faktorizace dají chápat jako plochy (od θ do 1) pod grafy podmíněných hustot $p(f|x)$ v likelihood diagramu (viz obrázky 4.1 a 4.2):

$$\begin{aligned} FPR(\theta) &= \int_{\theta}^1 p(f|x = 0) df, \\ TPR(\theta) &= \int_{\theta}^1 p(f|x = 1) df. \end{aligned} \quad (4.3)$$

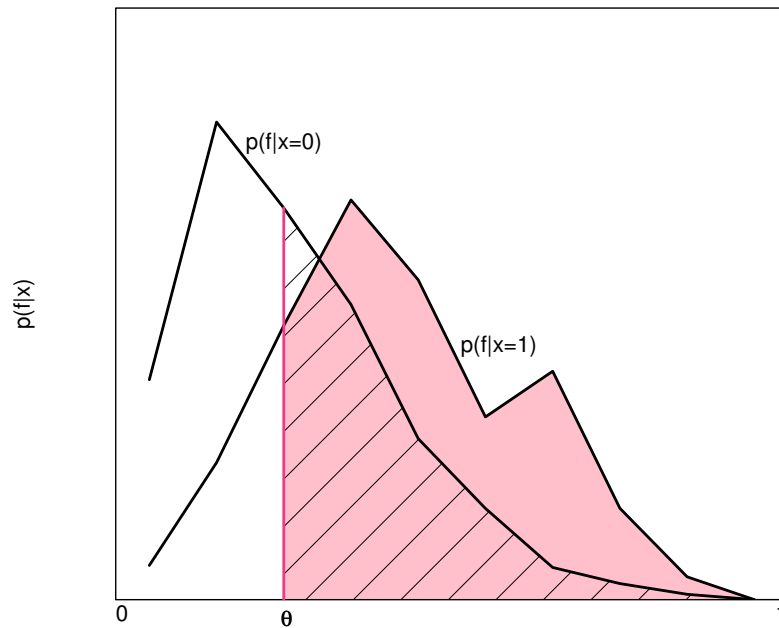
Plocha pod grafem ROC křivky se dá vyjádřit

$$AUC = \int_0^1 TPR(\theta) dFPR(\theta). \quad (4.4)$$

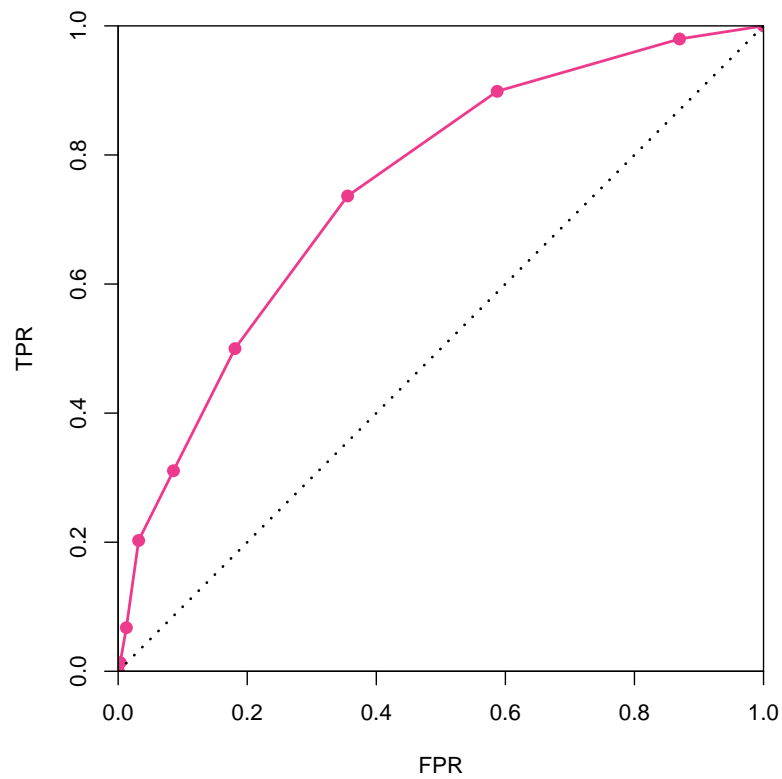
V likelihood diagramu tomu odpovídá

$$AUC = \int_0^1 \left[\int_{\theta}^1 p(f|x = 1) df \cdot p(\theta|x = 0) \right] d\theta, \quad (4.5)$$

kde výraz v hranaté závorce je v každém bodě θ rovný součinu obsahu plochy pod grafem funkce $p(f|x = 1)$ (od θ do 1) a hodnotě funkce $p(f|x = 0)$ v bodě θ (viz obrázek 4.1).



Obrázek 4.1: Likelihood diagram - $FPR(\theta)$ se rovná obsahu vyšrafované plochy, $TPR(\theta)$ se rovná obsahu barevné plochy, AUC je rovna integrálu přes všechna $\theta \in [0, 1]$ ze součinu obsahu barevné plochy a délky barevné úsečky.



Obrázek 4.2: ROC křivka spočítaná z likelihood diagramu pomocí vztahů (4.3).

Kapitola 5

Binormální model a MSE

5.1 Binormální model

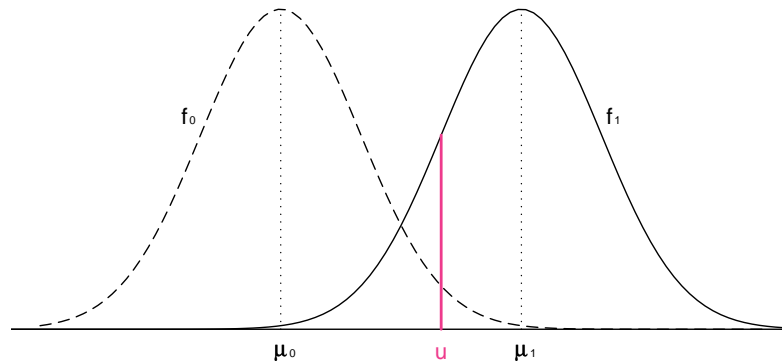
V binormálním modelu předpokládáme, že podmíněná rozdělení skórové funkce H na objektech z \mathcal{X}_1 , resp. z \mathcal{X}_0 , mají normální rozdělení se středními hodnotami μ_1 , resp. μ_0 , a rozptyly σ_1^2 , resp. σ_0^2 . Pro jednoduchost budeme brát $\sigma_0^2 = \sigma_1^2 = \sigma^2$ a $\mu_0 < \mu_1$. Viz obrázek 5.1.

Pravděpodobnostní předpověď F (skórovou funkci s hodnotami v $[0, 1]$) pak můžeme vytvořit následovně:

$$F(H(w)) = F(u) = \frac{f_1(u)}{f_0(u) + f_1(u)}, \quad (5.1)$$

kde f_1 , resp. f_0 , jsou podmíněné hustoty skórové funkce H na \mathcal{X}_1 , resp. \mathcal{X}_0 , a $u = H(w) \in \mathbf{R}$. Pravděpodobnostní předpověď F je pak rovna podmíněné pravděpodobnosti $P(X = 1|u)$. Vzorec (5.1) se v binormálním modelu (při $\sigma_0^2 = \sigma_1^2 = \sigma^2$ a $\mu_1 > \mu_0$) upraví na

$$F(u) = \frac{1}{1 + \exp\left\{-\left(\frac{\mu_0^2 - \mu_1^2}{2\sigma^2} + \frac{\mu_1 - \mu_0}{\sigma^2}u\right)\right\}}, \quad (5.2)$$



Obrázek 5.1: Binormální model – f_0 a f_1 jsou podmíněné hustoty rozdělení skórové funkce H na objektech z \mathcal{X}_0 a z \mathcal{X}_1 .

což je distribuční funkce logistického rozdělení s parametry $a = \frac{\mu_0^2 - \mu_1^2}{2\sigma^2}$ a $b = \frac{\mu_1 - \mu_0}{\sigma^2}$.
Důkaz:

$$\begin{aligned}
F(w) &= \frac{f_1(u)}{f_0(u) + f_1(u)} \\
&= \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(u-\mu_1)^2}{2\sigma^2}\right\}}{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(u-\mu_0)^2}{2\sigma^2}\right\} + \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(u-\mu_1)^2}{2\sigma^2}\right\}} \\
&= \frac{\exp\left\{-\frac{(u-\mu_1)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(u-\mu_1)^2}{2\sigma^2}\right\} \left(1 + \exp\left\{-\frac{(u-\mu_0)^2}{2\sigma^2} + \frac{(u-\mu_1)^2}{2\sigma^2}\right\}\right)} \\
&= \frac{1}{1 + \exp\left\{-\frac{-u^2 + 2\mu_0 u - \mu_0^2 + u^2 - 2\mu_1 u + \mu_1^2}{2\sigma^2}\right\}} \\
&= \frac{1}{1 + \exp\left\{\frac{\mu_1^2 - \mu_0^2 + 2(\mu_0 - \mu_1)u}{2\sigma^2}\right\}} \\
&= \frac{1}{1 + \exp\left\{-\left(\frac{\mu_0^2 - \mu_1^2}{2\sigma^2} + \frac{\mu_1 - \mu_0}{\sigma^2} u\right)\right\}}
\end{aligned}$$

□

5.2 MSE v binormálním modelu

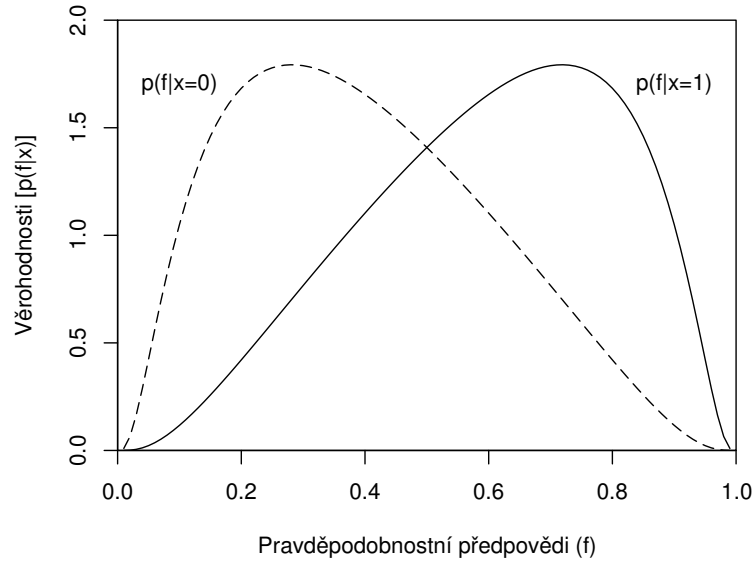
Definice MSE (3.12) se dá zobecnit i na případy, kdy mají předpovědi spojité rozdělení:

$$\text{MSE}(f, x) = \text{E}(f - x)^2 = \int_0^1 \sum_x (f - x)^2 p(f, x) df \quad (5.3)$$

V binormálním modelu, kde jsou pravděpodobnostní předpovědi definovány vztahy (5.1) a (5.2), můžeme spočítat podmíněné hustoty rozdělení předpovědí na objektech z \mathcal{X}_0 a z \mathcal{X}_1 . Platí

$$\begin{aligned}
p(f|x=0) &= \sqrt{\frac{\sigma^2}{2\pi}} \frac{1}{\mu_1 - \mu_0} \frac{1}{f(1-f)} \exp\left\{-\frac{\left(\frac{\sigma^2 \ln \frac{f}{1-f}}{\mu_1 - \mu_0} + \frac{\mu_1 - \mu_0}{2}\right)^2}{2\sigma^2}\right\}, \\
p(f|x=1) &= \sqrt{\frac{\sigma^2}{2\pi}} \frac{1}{\mu_1 - \mu_0} \frac{1}{f(1-f)} \exp\left\{-\frac{\left(\frac{\sigma^2 \ln \frac{f}{1-f}}{\mu_1 - \mu_0} + \frac{\mu_0 - \mu_1}{2}\right)^2}{2\sigma^2}\right\}.
\end{aligned} \quad (5.4)$$

Grafy podmíněných hustot rozdělení předpovědí na objektech z \mathcal{X}_0 a z \mathcal{X}_1 jsou na obrázku 5.2. Jedná se vlastně o likelihood diagram se spojitými křivkami vyjádřenými vzorci (5.4).



Obrázek 5.2: Likelihood diagram – podmíněné hustoty rozdělení předpovědi na objektech z \mathcal{X}_0 a z \mathcal{X}_1 (pro $\mu_1 - \mu_0 = 1$ a $\sigma^2 = 1$) – čárkovaná čára je $p(f|x=0)$, plná $p(f|x=1)$.

Důkaz: V binormálním modelu je podmíněné rozdělení H na objektech z \mathcal{X}_0 normální s hustotou

$$f_0(u) = f(u|x=0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(u-\mu_0)^2}{2\sigma^2}\right\}.$$

Pravděpodobnostní předpověď je

$$f = F(u) = \frac{1}{1 + \exp\{-(a + bu)\}},$$

kde $a = \frac{\mu_0^2 - \mu_1^2}{2\sigma^2}$ a $b = \frac{\mu_1 - \mu_0}{\sigma^2}$.

Použije se věta o transformaci náhodné veličiny. Funkce F je rostoucí na \mathbf{R} a jejím oborem hodnot je otevřený interval $(0, 1)$. Inverzní funkce k funkci F je

$$u = G(f) = \frac{\ln \frac{f}{1-f} - a}{b}.$$

Derivace funkce G v bodě f je

$$\begin{aligned} G'(f) &= \frac{1}{b} \left(\frac{1-f}{f} \frac{(1-f) + f}{(1-f)^2} \right) \\ &= \frac{1}{b(1-f)f}. \end{aligned}$$

Derivace je kladná, neboť $f \in (0, 1)$ a $b > 0$.

Podle věty o transformaci náhodné veličiny je pak hustota rozdělení předpo-

vědí na \mathcal{X}_0 rovna

$$\begin{aligned}
 p(f|x=0) &= f_0 \left(\frac{\ln \frac{f}{1-f} - a}{b} \right) |G'(f)| \\
 &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{\left(\frac{\ln \frac{f}{1-f} - a}{b} - \mu_0 \right)^2}{2\sigma^2} \right\} \frac{1}{b(1-f)f} \\
 &= \sqrt{\frac{\sigma^2}{2\pi}} \frac{1}{\mu_1 - \mu_0} \frac{1}{f(1-f)} \exp \left\{ -\frac{\left(\frac{\sigma^2 \ln \frac{f}{1-f}}{\mu_1 - \mu_0} + \frac{\mu_1 - \mu_0}{2} \right)^2}{2\sigma^2} \right\}.
 \end{aligned}$$

Důkaz pro $p(f|x=1)$ se provede analogicky.

□

Známe-li rozdělení pozorování $p(x)$, můžeme MSE vyjádřit následovně:

$$\text{MSE}(f, x) = \int_0^1 \sum_{x=0}^1 (f - x)^2 p(f|x) p(x) df \quad (5.5)$$

Do tohoto vztahu se dají dosadit podmíněné pravděpodobnosti z rovností (5.4).

Kapitola 6

Aplikace na konkrétní předpovědi

Diagnostický přístup k ověřování pravděpodobnostních předpovědí ilustruji na datech získaných v rámci výzkumu „Aktéři, rizika a důvěra“. Tento výzkum byl proveden v prosinci roku 2005 na katedře sociologie Filosofické fakulty Univerzity Karlovy pod vedením Doc. PhDr. Jiřího Buriánka. Byl dílčím úkolem projektu „Rozvoj české společnosti v EU: výzvy a rizika“, který probíhá v letech 2005 – 2010 a jehož koordinátorem je profesor Jiří Kabele.

Výzkum sledoval, jak bezpečně se lidé cítí večer po setmění a v noci, když někam jdou v okolí svého bydliště. Na základě informací o respondentech (věk, pohlaví, vzdělání, lidnatost obce, vztah k drogám, ...) byly pomocí logistické regrese vytvořeny pravděpodobnostní předpovědi toho, jestli se daný člověk „bojí“.

6.1 Sdružené rozdělení

Pravděpodobnostní předpovědi získané výzkumem jsem zaokrouhlila na hodnoty z množiny $\{0.05, 0.15, 0.25, \dots, 0.95\}$, aby se s nimi dalo pracovat jako s diskretní náhodnou veličinou. Skutečná pozorování nabývají dvou různých hodnot – buď se člověk bojí ($x = 1$) nebo se nebojí ($x = 0$).

Základní míry sdruženého rozdělení $p(f, x)$ jsou v tabulce 6.1. Korelační koeficient dává představu o závislosti mezi f a x . Čím větší by bylo ρ_{fx} , tím silněji by na sobě předpovědi a pozorování závisely.

Tabulka 6.1: Základní míry sdruženého rozdělení a marginálních rozdělení předpovědi a pozorování.

Střední hodnoty		Rozptyly		Korelační koeficient	Velikost vzorku
μ_f	μ_x	σ_f^2	σ_x^2	ρ_{fx}	n
0.313	0.320	0.033	0.217	0.409	463

Tabulka 6.2: Základní míry podmíněného rozdělení předpovědí.

Střední hodnoty		Rozptyly		Velikosti vzorků	
$\mu_{f x=0}$	$\mu_{f x=1}$	$\sigma_{f x=0}^2$	$\sigma_{f x=1}^2$	$n(x=0)$	$n(x=1)$
0.263	0.421	0.025	0.032	315	148

6.2 CR faktorizace

Diagram vlastností sdruženého rozdělení předpovědí a pozorování, které vyplývají z CR faktorizace, se nazývá attributes diagram (obrázek 6.1). Skládá se ze dvou částí. Ve spodní části, která se nazývá kalibrační diagram, je graf $\mu_{x|f}$. Horní část je graf rozdělení předpovědí $p(f)$.

Poznámka: Někdy se výrazem attributes diagram myslí jen kalibrační diagram (spodní část obrázku 6.1).

Kalibrační diagram (obrázek 6.1 dole) ukazuje vztah mezi $\mu_{x|f}$ a f a obsahuje i několik pomocných čar. Čárkovaná čára (diagonála), pro niž $\mu_{x|f} = f$, reprezentuje perfektně kalibrované předpovědi. Vodorovná čerchovaná čára, představuje naprosto nerozlišující předpovědi, při nichž $\mu_{x|f} = \mu_x$ pro všechny $f \in F$. Tečkovaná čára, která je stejně vzdálená od diagonály i od vodorovné úsečky, reprezentuje předpovědi s nulovou úspěšností ($SS = 0$). Plyne to z definice SS a z CR faktorizace střední čtvercové chyby. Úspěšnost je kladná, pokud je MSE pro danou předpověď větší než MSE konstantní předpovědi rovnající se μ_x . Na obrázku 6.2 je úspěšnost kladná v barevných částech kalibračního diagramu, v bílých je záporná.

Z grafu rozdělení $p(f)$ (obrázek 6.1 nahoře) je vidět, že malé hodnoty předpovědí a hodnoty blízko base rate jsou používány mnohem častěji než velké pravděpodobnosti.

6.3 LBR faktorizace

Likelihood diagram (obrázek 6.3) obsahuje dvě křivky (plnou čáru $p(f|x=0)$ a čárkovanou $p(f|x=1)$). Tečkovaná čára je střední hodnota pozorování μ_x .

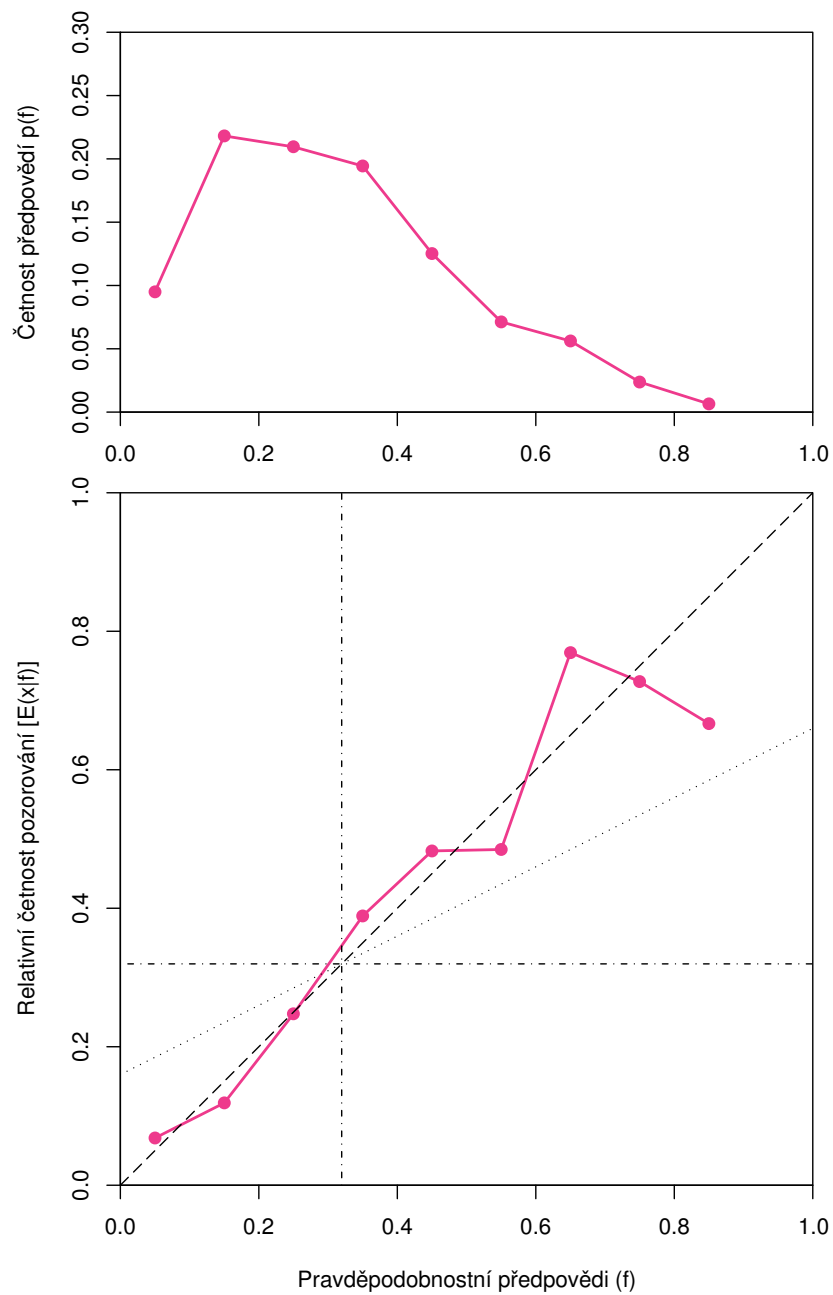
V ideálním případě by se oblasti pod křivkami $p(f|x=0)$ a $p(f|x=1)$ neměly moc překrývat. V případě, kdy $x=0$, by se měly vyskytovat jen nízké hodnoty pravděpodobnostních předpovědí a při $x=1$ by měly být předpovědi velké.

Za povšimnutí také stojí, že bod, kde se křivky $p(f|x=0)$ a $p(f|x=1)$ protínají, se nachází v blízkosti střední hodnoty pozorování μ_x (ta je v obrázku 6.3 vyznačená tečkovanou čarou).

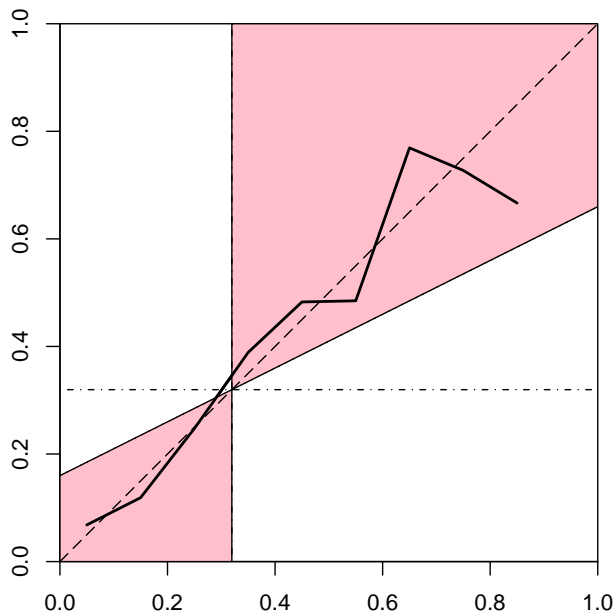
Základní míry podmíněného rozdělení jsou v tabulce 6.2

6.4 Míry kvality předpovědí

Střední chyba, střední čtvercová chyba a skill score jsou v tabulce 6.3.



Obrázek 6.1: Attributes diagram – rozdělení předpovědí (nahore) a kalibrační diagram (dole): čárkovaná čára (diagonála) představuje perfektně kalibrované předpovědi, čerchované čáry nerozlišující (no resolution) předpovědi a tečkovaná čára předpovědi s nulovým skill score.



Obrázek 6.2: Kalibrační diagram – úspěšnost je kladná v barevných částech, jinak je záporná.

Tabulka 6.3: Performance measures.

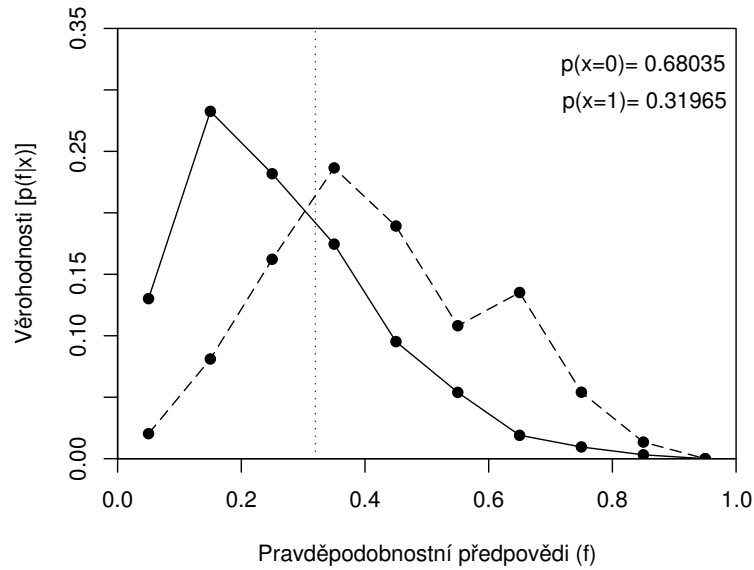
ME nepodmíněné vychýlení	MSE celková správnost	SS (%) úspěšnost
-0.00637	0.18129	16.6%

Pro porovnávání několika předpovědních systémů se srovnávají hodnoty ME, MSE a SS. Čím jsou ME a MSE menší, tím jsou předpovědi méně vychýlené a přesnější. S rostoucím SS úspěšnost předpovědí zvětšuje.

6.5 Rozklady MSE

Základní rozklad MSE

Základní rozklad MSE je v tabulce 6.4. Na první pohled je z tabulky vidět, že vliv vychýlení na MSE je v tomto případě zanedbatelný. Rozptyl pozorování, který nemůže prognostik nijak ovlivnit, je větší než rozptyl předpovědí, a přispívá tedy ke zvětšení MSE (podstatně) víc než rozptyl předpovědí. Kovariance předpovědí a pozorování ($2\sigma_f\sigma_x\rho_{fx}$) MSE zmenšuje.



Obrázek 6.3: Likelihood diagram – plná čára je $p(f|x = 0)$, čárkovaná $p(f|x = 1)$ a tečkovaná je střední hodnota pozorování μ_x .

Tabulka 6.4: Základní rozklad MSE.

MSE celková správnost	=	$(\mu_f - \mu_x)^2$ vychýlení	+	σ_f^2	+	σ_x^2	-	$2\sigma_f\sigma_x\rho_{fx}$
								rozptyl chyb předpovědí
0.18129		0.00004		0.03260		0.21748		0.06883

CR rozklad MSE

Z CR rozkladu MSE (tabulka 6.5) je vidět, že přírůstek MSE způsobený nekalibrováností (druhý člen rozkladu) je minimální. Zlepšení založené na rozlišení (třetí člen rozkladu) překoná zhoršení z nekalibrovánosti, a úspěšnost (skill) je tedy v tomto případě kladná.

LBR rozklad MSE

LBR rozklad MSE je v tabulce 6.6. Druhý a třetí člen rozkladu (dohromady) udávají „pozici“ podmíněné střední hodnoty předpovědí při jednotlivých pozorováních $\mu_{f|x}$ a nepodmíněné střední hodnotě předpovědí μ_f .

Tabulka 6.5: CR rozklad MSE.

MSE celková správnost	=	σ_x^2 nejistota	+	$E_f(\mu_{x f} - f)^2$ kalibrace	-	$E_f(\mu_{x f} - \mu_x)^2$ rozlišení
0.18129		0.21748		0.00200		0.03819

Tabulka 6.6: LBR rozklad MSE.

MSE celková správnost	=	σ_f^2 ostrota	+	$E_x(\mu_{f x} - x)^2$ podmíněné vychýlení	-	$E_x(\mu_{f x} - \mu_f)^2$ diskriminace
0.18129		0.03260		0.15413		0.00545

Tabulka 6.7: Základní rozklad SS.

SS celková správnost	=	ρ_{fx}^2 sdružení	-	$[\rho_{fx} - (\sigma_f/\sigma_x)]^2$ kalibrace	-	$[(\mu_f - \mu_x)/\sigma_x]^2$ vychýlení
0.16639		0.16704		0.00046		0.00019

6.6 Rozklad SS

Výsledky rozkladu SS jsou v tabulce 6.7. Je vidět, že pro tyto předpovědi mají nekalibrovanost a vychýlení (druhý a třetí člen rozkladu) jen minimální vliv na velikost SS.

6.7 ROC křivka

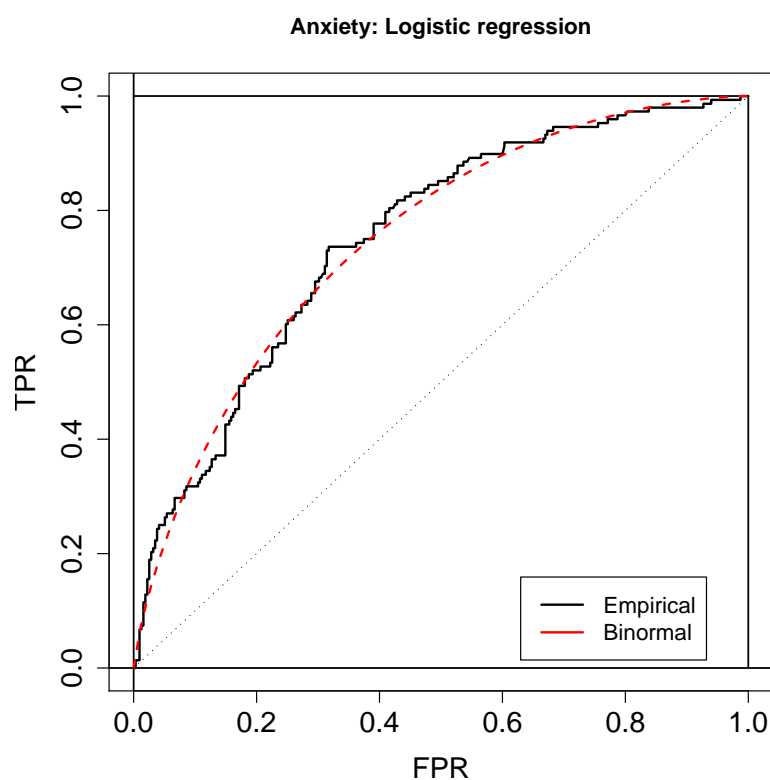
Na obrázku 6.4 jsou dvě ROC křivky. Barevná křivka je spočítaná z binornálního modelu (viz kapitola 5).

Graf ROC křivky, který odpovídá LBR faktorizaci (obrázek 6.3) a který je spočítaný pomocí vzorců (4.3), je na obrázku 4.2.

Případný rozdíl mezi ROC křivkami na obrázcích 6.4 a 4.2 je způsoben tím, že předpovědi, ze kterých je vytvořen likelihood diagram (obrázek 6.3) a následně pak ROC křivka na obrázku 4.2, vznikly „zaokrouhlením“ skutečných předpovědí na hodnoty z množiny $\{0.05, 0.15, \dots, 0.95\}$. Graf 6.4 je vytvořen přímo z nezaokrouhlených předpovědí.

6.8 Shrnutí

1. Nepodmíněné vychýlení je zanedbatelné, předpovědi jsou prakticky nevychýlené (viz tabulky 6.1, 6.3, 6.4, 6.7).
2. Předpovědi jsou dobře kalibrované (viz obrázek 6.1, tabulka 6.7).
3. Předpovědi nejsou příliš ostré, vysoké pravděpodobnostní hodnoty předpovědí se používají velmi málo (viz obrázek 6.1, tabulky 6.1, 6.6).
4. Předpovědi vykazují kladnou úspěšnost (tj. kladný skill score), která je především ovlivněna hodnotou korelace mezi předpověďmi a pozorováními (viz tabulka 6.7).



Obrázek 6.4: ROC křivka - převzato od Mgr. Martina Betince.

5. Podmíněné střední hodnoty předpovědí $\mu_{f|x=1}$ a $\mu_{f|x=0}$ jsou blíže k nepodmíněné střední hodnotě předpovědí μ_f než ke skutečným pozorováním (viz tabulky 6.1, 6.2, 6.6).

Celkově hlavní nedostatek předpovědí je v malé ostrosti, hlavními přednostmi jsou dobrá kalibrování a nevychýlenost.

Informace obsažené v ROC křivce jsou stejné jako informace vyplývající z LBR faktorizace. Není proto příliš velkým přínosem tuto křivku konstruovat.

Literatura

- [1] Betinec, M. (2006). Použití ROC křivek pro hodnocení klasifikátorů. Ve: *Sborník prací 14. zimní školy JČMF ROBUST 2006*, eds. J. Antoch & G. Dohnal, 25-34, JČMF.
- [2] Doolittle, M. H. (1885). The verification of predictions. *American Meteorological Journal*, **2**, 327-329.
- [3] Finley, J. P. (1884). Tornado prediction. *American Meteorological Journal*, **1**, 85-88.
- [4] Gilbert, G. K. (1884). Finley's tornado predictions. *American Meteorological Journal*, **1**, 166-172.
- [5] Murphy, A. H. (1996). The Finley Affair: A Signal Event in the History of Forecast Verification. *Weather and Forecasting*, **11**, 3-20.
- [6] Murphy, A. H. (1997). Forecast Verification. In: *Economic Value of Weather and Climate Forecasts*, eds. R. W. Katz & A. H. Murphy, 19-74, Cambridge University Press.
- [7] Murphy, A. H. & Winkler, R. L. (1992). Diagnostic verification of probability forecasts. *International Journal of Forecasting*, **7**, 435-455.
- [8] Peirce, C. S. (1884). The numerical measure of the success of predictions. *Science*, **4**, 453-454.