

# Sannolikhet och statistik XI

Johan Jonasson

April 2019

**Betingade väntevärden.** Vi ska säga att  $\mathbb{E}[Y|X = x]$  är väntevärdet av den sv som samma förd som  $Y$  givet  $X = x$ .

*Definition:*

- $Y$  diskret:  $\mathbb{E}[Y|X = x] = \sum_{y_k \in V_Y} y_k p_{Y|X}(y_k|x),$
- $Y$  kont:  $\mathbb{E}[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy.$

*Totala sannolikhetslagen för väntevärden:*

- $X$  diskret:  $\mathbb{E}[Y] = \sum_{x_j \in V_X} \mathbb{E}[Y|X = x_j] p_X(x_j),$
- $X$  kont:  $\mathbb{E}[Y] = \int_{-\infty}^{\infty} E[Y|X = x] f_X(x) dx.$

Bevis: Fallet  $(X, Y)$  kont.

$$\begin{aligned} \int_{-\infty}^{\infty} \mathbb{E}[Y|X = x] f_X(x) dx &= \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy \right) f_X(x) dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y) dy dx \\ &= \mathbb{E}[Y] \end{aligned}$$

där sista likheten är OSL.

Exempel forts:  $X \sim \text{likf}(0, 1)$  och sedan  $Y \sim \text{likf}(0, X)$ . Vi såg förut att  $f(x, y) = 1/x$ ,  $0 < y < x < 1$ . Enl OSL

$$\mathbb{E}[Y] = \int_0^1 \int_0^x y f(x, y) dy dx = \int_0^1 \int_0^x y \frac{1}{x} dy dx = \frac{1}{2} \int_0^1 \frac{1}{x} x^2 dx = \frac{1}{4}.$$

Alt.

$$\mathbb{E}[Y] = \int_0^1 \mathbb{E}[Y|X = x] f_X(x) dx = \int_0^1 \frac{1}{2} x dx = \frac{1}{4}.$$

*Definition:*  $\mathbb{E}[Y|X]$  är den sv som antar värdet  $\mathbb{E}[Y|X = x]$  då  $X = x$ .

M.a.o. om fknen  $g$  ges av  $g(x) = \mathbb{E}[Y|X = x]$  så är  $\mathbb{E}[Y|X] = g(X)$ .

Kortform av TSL för bet vv: Antag  $(X, Y)$  kont (funkar lika bra diskret). Enl TSL och OSL gäller

$$\mathbb{E}[Y] = \int \mathbb{E}[Y|X = x]f_X(x)dx = \int g(x)f_X(x)dx = \mathbb{E}[g(X)] = \mathbb{E}[\mathbb{E}[Y|X]].$$

Detta är värt highlight:

$$\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y].$$

*Exempel:* Låt  $X \sim \text{Geo}(p)$  och sedan  $Y \sim \text{Bin}(X, r)$ . Vad är  $\mathbb{E}[Y]$ ?

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Xr] = \frac{r}{p}.$$

□

*Exempel.* Låt  $X_1, X_2, \dots$  vara oberoende och  $\mathbb{E}[X_k] = \mu$  för alla  $k$ .  
Låt  $N$  vara ober av  $X_k$ :na och pos. heltalsvärd.

$$\mathbb{E} \left[ \sum_{k=1}^N X_k \right] = \mathbb{E} \left[ \mathbb{E} \left[ \sum_{k=1}^N X_k \mid N \right] \right] = \mathbb{E}[N\mu] = \mu\mathbb{E}[N].$$

□

Obs att

$$\mathbb{P}(A) = \mathbb{E}[\mathbb{P}(A|X)]$$

$$\text{ty h.l.} = \int \mathbb{P}(A|X = x) f_X(x) dx = \mathbb{P}(A).$$

**Kovarians och korrelation.** Låt  $X$  och  $Y$  vara två sv.

*Definition:*  $\text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$ .

Ett mått på hur  $X$  och  $Y$  samvarierar:  $\text{Cov}(X, Y) > 0$  betyder positivt beroende och  $\text{Cov}(X, Y) < 0$  negativt beroende.

Formel:  $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ .

Följer av att

$$\text{Cov}(X, Y) = \mathbb{E}[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y] = \mathbb{E}[XY] - \mu_X \mu_Y.$$

Om  $X, Y$  oberoende gäller  $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ , så formeln ger att

$$X \text{ och } Y \text{ oberoende} \Rightarrow \text{Cov}(X, Y) = 0.$$

Varning:  $\text{Cov}(X, Y) = 0$  medför inte i allmänhet att  $X$  och  $Y$  oberoende. Motexempel strax.

Några obs:

- $\text{Cov}(X, a) = 0,$
- $\text{Cov}(X, X) = \text{Var}(X),$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X).$

Formeln ger också att kovariansen är bilinjär:

$$\begin{aligned} \text{Cov}(a_1 X_1 + a_2 X_2, b_1 Y_1 + b_2 Y_2) &= a_1 b_1 \text{Cov}(X_1, Y_1) + a_1 b_2 \text{Cov}(X_1, Y_2) \\ &+ a_2 b_1 \text{Cov}(X_2, Y_1) + a_2 b_2 \text{Cov}(X_2, Y_2). \end{aligned}$$

Exempelvis  $\text{Cov}(X + a, Y + b) =$   
 $\text{Cov}(X, Y) + \text{Cov}(X, b) + \text{Cov}(a, Y) + \text{Cov}(a, b) = \text{Cov}(X, Y).$

Variansen för en summa:

$$\begin{aligned}\text{Var}(X + Y) &= \text{Cov}(X + Y, X + Y) \\ &= \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y).\end{aligned}$$

*Exempel:* Låt  $\Theta \sim \text{likf}(0, 2\pi)$ , och  $X = \cos \Theta$ ,  $Y = \sin \Theta$ . Vi har

$$\mathbb{E}[X] = \int_0^{2\pi} \cos \theta \frac{1}{2\pi} d\theta = 0, \quad \mathbb{E}[Y] = \int_0^{2\pi} \sin \theta \frac{1}{2\pi} d\theta = 0.$$

$$\mathbb{E}[XY] = \int_0^{2\pi} \cos \theta \sin \theta \frac{1}{2\pi} d\theta = 0.$$

Alltså:  $\text{Cov}(X, Y) = 0 - 0 \cdot 0 = 0$ , men  $X$  och  $Y$  är inte oberoende. □



*Exempel:* Låt  $X$  och  $Y$  vara oberoende och *likf*(0, 1). Låt  $Z = XY$ ,  $W = X + Y$ . Bestäm  $\text{Cov}(Z, W)$ .

$$\mathbb{E}[Z] = \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = \frac{1}{4}, \quad \mathbb{E}[W] = \mathbb{E}[X] + \mathbb{E}[Y] = \frac{1}{2}.$$

$$\mathbb{E}[ZW] = \mathbb{E}[XY(X + Y)] = \mathbb{E}[X^2Y] + \mathbb{E}[XY^2] = \mathbb{E}[X^2]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y^2] = \frac{1}{3}.$$

Alltså

$$\text{Cov}(Z, W) = \frac{1}{3} - \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{12}.$$

Alltså positivt beroende. □

OK, så kovariansen var  $1/12$  i förra exemplet, men vad säger det om hur *starkt* det positiva beroendet är? Måste sättas i relation till varianserna. Dags att prata om korrelation.

*Schwarz olikhet:* För alla sv  $X$  och  $Y$  gäller

$$\mathbb{E}[XY]^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2].$$

*Bevis:* Vi anv att för alla konstanter  $t$ :

$$0 \leq \mathbb{E}[(X - tY)^2] = \mathbb{E}[X^2] - 2t\mathbb{E}[XY] + t^2\mathbb{E}[Y^2].$$

H.l. blir som minst då  $t = \mathbb{E}[XY]/\mathbb{E}[Y^2]$ . Sätt in detta  $t$  i h.l. och få

$$0 \leq \mathbb{E}[X^2] - 2\frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]} + \frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]} = \mathbb{E}[X^2] - \frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]}.$$

Förläng med  $\mathbb{E}[Y^2]$  och saken är klar. □

*Definition:* Korrelationskoefficienten för  $(X, Y)$  ges av

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

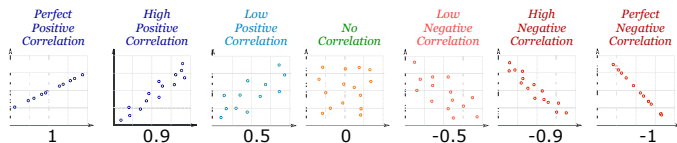
Vi har

$$-1 \leq \rho(X, Y) \leq 1$$

enligt Schwarz anv på  $X - \mu_X$  och  $Y - \mu_Y$ . Vi har också

- $\rho(X, Y) = \rho(aX, bY)$  (skalningsinvarians),
- $Y = aX + b$  för konstant  $a > 0 \Leftrightarrow \rho(X, Y) = 1$ ,
- $Y = aX + b$  för konstant  $a < 0 \Leftrightarrow \rho(X, Y) = -1$ .

Typiska plottar av data  $(X_1, Y_1), \dots, (X_n, Y_n)$ :



Figur: Korrelationskoefficient.

*Definition:* Om  $\rho(X, Y) = 0$  kallas  $X$  och  $Y$  okorrelerade.

Exempel forts: Vi hade  $X, Y$  oberoende och  $\text{likf}(0, 1)$ , och  $Z = XY$ ,  $W = X + Y$ . Vi fick  $\text{Cov}(Z, W) = 1/12$ . Vad är  $\rho(X, Y)$ ? Vi behöver varianserna.

$$\text{Var}(W) = \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) = \frac{1}{6}$$

eftersom variansen av  $\text{likf}(0, 1)$  är  $1/12$ .

$$\begin{aligned} \text{Var}(Z) &= \text{Var}(XY) = \mathbb{E}[X^2Y^2] - \mathbb{E}[XY]^2 \\ &= \mathbb{E}[X^2]\mathbb{E}[Y^2] - (\mathbb{E}[X]\mathbb{E}[Y])^2 \\ &= \left(\frac{1}{3}\right)^2 - \left(\frac{1}{4}\right)^2 = \frac{7}{144} \end{aligned}$$

eftersom  $\mathbb{E}[X^2] = 1/3$ . Alltså

$$\rho(Z, W) = \frac{1/12}{\sqrt{(1/6) \cdot (7/144)}} = \sqrt{\frac{6}{7}}.$$



*Exempel:* Bästa linjära prediktor: Välj  $a$  och  $b$  för att minimera  $\mathbb{E}[(Y - (aX + b))^2]$ .

Antag att  $\mathbb{E}[X] = \mathbb{E}[Y] = 0$  och  $\text{Var}(X) = \text{Var}(Y) = 1$ . Då är  $\rho = \text{Cov}(X, Y) = \mathbb{E}[XY]$  och

$$\begin{aligned}\mathbb{E}[(Y - (aX + b))^2] &= \mathbb{E}[Y^2] - 2\mathbb{E}[Y(aX + b)] + \mathbb{E}[(aX + b)^2] \\ &= 1 - 2a\rho + a^2 + b^2.\end{aligned}$$

Detta min med  $b = 0$  och  $a = \rho$ . Man säger alltså att i detta fall är bästa linjära prediktor av  $Y$  givet  $X$  lika med  $\rho X$ . I det allmänna fallet ges den bästa linjära prediktorn av

$$\mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (X - \mu_X).$$

□