# TAMS79: Föreläsning 7
# Betingning och kovarians

Johan Thim (johan.thim@liu.se)

10 november 2018

## 7.1 Betingade fördelningar

Om vi har en flerdimensionell stokastisk variabel kan det vara intressant att fundera på hur saker hänger ihop om vi vet utfallet av en komponent. För enkelhetens skull begränsar vi oss till två dimensioner. Vi definierar den betingade sannolikhets- respektive täthetsfunktionen enligt följande.

**Betingad fördelning**

**Definition.** Låt $(X, Y)$ ha sannolikhetsfunktionen $p_{X,Y}(x, y)$. Då definierar vi sannolikhetsfunktionen för $X$ givet att $Y = y$ enligt

$$p_{X|Y=y}(x\,;\,y) = \frac{p_{X,Y}(x, y)}{p_Y(y)},$$

under förutsättning att $p_Y(y) > 0$. På motsvarande sätt definierar vi – om $(X, Y)$ har den simultana täthetsfunktionen $f_{X,Y}(x, y)$ – den betingade täthetsfunktionen för $X$ givet $Y = y$ enligt

$$f_{X|Y=y}(x\,;\,y) = \frac{f_{X,Y}(x, y)}{f_Y(y)},$$

under förutsättning att $f_Y(y) > 0$. Motsvarande definitioner gäller för $Y \,|\, X = x$.

Det är en ganska naturlig definition om vi tänker tillbaka till hur vi definierade betingad sannolikhet, då $P(A \,|\, B) = P(A \cap B)/P(B)$. Notera även vad som händer om variablerna är oberoende:

$$f_{X|y=y}(x\,;\,y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{f_X(x) f_Y(y)}{f_Y(y)} = f_X(x).$$

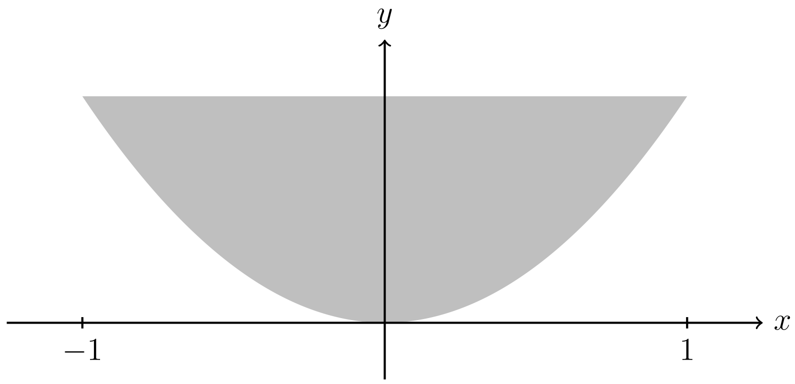
**Exempel**

Låt $(X,Y)$ vara likformigt fördelad på området $-1 < x < 1$, $x^2 < y < 1$. Hitta de betingade täthetsfunktionerna $f_{X\,|\,Y=y}(x)$ och $f_{Y\,|\,X=x}(y)$. Beräkna $P(Y > 0.5\,|\,X = x)$.

**Lösning**. Arean av området $D = \{(x,y) \in \mathbf{R}^2 : -1 \leq x \leq 1,\ x^2 \leq y \leq 1\}$ ges av

$$A = \int_{-1}^{1}\int_{x^2}^{1} dy\,dx = \int_{-1}^{1}(1-x^2)dx = 2 - \frac{2}{3} = \frac{4}{3},$$

så $f(x,y) = \dfrac{3}{4}$ för $(x,y) \in D$ och $f(x,y) = 0$ annars.




Vi kan beräkna de marginella tätheterna:

$$f_X(x) = \int_{x^2}^{1} \frac{3}{4}dy = \frac{3}{4}(1-x^2), \quad -1 < x < 1,$$

och

$$f_Y(y) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{3}{4}dx = \frac{3}{2}\sqrt{y}, \quad 0 < y < 1.$$

Alltså har vi de betingade tätheterna

$$f_{X\,|\,Y=y}(x) = \frac{3/4}{3\sqrt{y}/2} = \frac{1}{2\sqrt{y}}, \quad -\sqrt{y} < x < \sqrt{y},$$

för $0 < y < 1$ och

$$f_{Y\,|\,X=x}(y) = \frac{3/4}{3(1-x^2)/4} = \frac{1}{1-x^2}, \quad x^2 < y < 1,$$

för $-1 < x < 1$. Notera att de betingade tätheterna är likformigt fördelade på olika linjesegment. Vi kan nu beräkna att

$$P(Y > 0.5\,|\,X = x) = \int_{0.5}^{1} f_{Y\,|\,X=x}(y)\,dy = \frac{1}{1-x^2}\int_{\max\{x^2,0.5\}}^{1} dy = \frac{1-\max\{x^2,0.5\}}{1-x^2}.$$

### 7.1.1 Betingat väntevärde

Nu ska vi göra något skojjigt. Sug på den här formeln: $E(X) = E(E(X\,|\,Y))$. Vi kan alltså i någon mening räkna ut väntevärden ”partiellt,” på liknande sätt som vi kunde använda lagen om total sannolikhet för att förenkla livet tidigare. Vi definierar först vad vi menar med ett betingad väntevärde.

### Betingat väntevärde

**Definition.** Det betingade väntevärdet definieras enligt

$$E(X \mid Y = y) = \sum_k k p_{X|Y=y}(k)$$

i det diskreta fallet och

$$E(X \mid Y = y) = \int_{-\infty}^{\infty} x f_{X|Y=y}(x)\, dx$$

i det kontinuerliga fallet (givet att täthetsfunktionerna existerar).

En rakt på definition där vi explicit utnyttjar den betingade sannolikhets- eller täthetsfunktionen. Med denna definition i bagaget kan vi formulera en mycket användbar formel.

### Lagen om totalt väntevärde

**Sats.**

$$E(E(X \mid Y)) = E(X).$$

Denna sats förtjänar en liten kommentar om vad vi egentligen menar med tanke på den vid första anblick konstiga uppsynen. Det är klart att $g(y) = E(X \mid Y = y)$ definierar en funktion av $y$. För den stokastiska motsvarigheten $g(Y)$ kan vi försöka räkna ut väntevärdet:

$$E(g(Y)) = E(E(X \mid Y)).$$

Beviset är krångligare i det generella fallet (med kontinuerliga variabler) och kräver en hel del vi inte har tillgång till. Men i det diskreta fallet kan vi skissa hur det ser ut.
Formellt så gäller att

$$\begin{aligned} E(E(X \mid Y)) &= E\left(\sum_k k p_{X \mid Y}(k)\right) = \sum_m \left(\sum_k k p_{X \mid Y}(k)\right) p_Y(m) \\ &= \sum_m \sum_k k p_{X \mid Y}(k) p_Y(m) = \sum_m \sum_k k p_{X,Y}(k,m) \\ &= \sum_k k \sum_m p_{X,Y}(k,m) = \sum_k k p_X(k) = E(X), \end{aligned}$$

då under förutsättning att vi kan iterera summorna på detta sätt. Om allt är absolutkonvergent är detta ok.

### Exempel

Elizabeth är ute och promenerar i Australienska vildmarken och blir ordentligt biten av en orm. Ormen slingrar snabbt iväg, men Elizabeth är säker på att det är en inlandstaipan eller en mer vanlig brunorm. Den förväntade tiden till kroppen ger upp pga giften är ca 60 minuter för Taipanen och 24 timmar för brunormen. Elizabeth uppskattar sannolikheten att det var en Taipan till 1/10 då den är mindre vanligt. Beräkna den förväntade tiden till döden inträffar.

**Lösning**. Låt $Y = 0$ om ormen är en Taipan och $Y = 1$ om ormen är en brunorm. Låt $X$ vara tiden till döden inträffar. Då är $E(X \,|\, Y = 0) = 60$ minuter och $E(X \,|\, Y = 1) = 24 \cdot 60$ minuter. Vidare är $P(Y = 0) = 1/10$ och $P(Y = 1) = 9/10$. Enligt lagen om totalt väntevärde så är

$$\begin{aligned} E(X) &= E(E(X \,|\, Y)) = \sum_{k=0}^{1} kE(X \,|\, Y = k) \\ &= E(X \,|\, Y = 0)P(Y = 0) + E(X \,|\, Y = 1)P(Y = 1) = \frac{60}{10} + \frac{9 \cdot 60 \cdot 24}{10} = 1302. \end{aligned}$$

Elizabeth har alltså en förväntad livslängd på 21.7 timmar till.

### Exempel

Lena och Sture har simmat vilse i ett undervattensgrottsystem. De befinner sig i en större kammare med två vägar ut. Den ena leder ut till en grotta där de kommer ut ur grottsystemet efter 20 minuter. Den andra går i en labyrintliknande mardröm som tar 50 minuter att traversera innan man plötsligt befinner sig tillbaka i kammaren. Båda lider av koldioxidförgiftningen och beslut fattas på dåliga grunder, så man väljer vägen ut med sannolikheten 0.3 och labyrinten med sannolikhet 0.7. Man har luft kvar i gastuberna för 2 timmar. Räcker luften för den förväntade tiden att de kommer ut?

**Lösning**. Låt $Y = 0$ om man väljer labyrinten och $Y = 1$ om man väljer vägen ut. Definiera $X$ som antalet minuter innan man kommer ut ur grottsystemet. Enligt lagen om totalt väntevärde så gäller att

$$E(X) = E(X \,|\, Y = 0)P(Y = 0) + E(X \,|\, Y = 1)P(Y = 1) = (E(X) + 50) \cdot 0.7 + 20 \cdot 0.3.$$

Om vi väljer labyrintvägen så kommer vi tillbaka och får helt enkelt börja om. Det innebär att den förväntade tiden till att man tar sig ut helt enkelt blir 50 minuter längre. Ur ekvationen ovan kan vi lösa ut $E(X)$:

$$(1 - 0.7)E(X) = 50 \cdot 0.7 + 20 \cdot 0.3 = 41 \quad \Leftrightarrow \quad E(X) = 136.7.$$

Det verkar alltså inte lovande.

Betingad varians definieras analogt med betingat väntevärde. Det finns även en formel som brukar kallas för lagen om total varians.

### Lagen om total varians

**Sats.** Om $V(Y) < \infty$ gäller att

$$V(Y) = E(V(Y \,|\, X)) + V(E(Y \,|\, X)),$$

om $X$ och $Y$ har samma underliggande sannolikhetsrum.

**Bevis**. Vi betingar på $X$ och utnyttjar lagen om totalt väntevärde:

$$\begin{aligned} V(Y) &= E(Y^2) - E(Y)^2 = E(E(Y^2 \,|\, X)) - E(E(Y \,|\, X))^2 \\ &= E(V(Y \,|\, X) + E(Y \,|\, X)^2) - E(E(Y \,|\, X))^2 \\ &= E(V(Y \,|\, X)) + E(E(Y \,|\, X)^2) - E(E(Y \,|\, X))^2 = E(V(Y \,|\, X)) + V(E(Y \,|\, X)). \end{aligned}$$

## 7.2 Kovarians

Vi har betraktat variansen för stokastiska variabler, men kan man säga någon om variationen mellan två variabler utan att stirra sig blind på sannolikhets- eller täthetsfunktionen? Visst kan man det, och svaret kommer i form av kovarians eller korellation.

**Definition.** Låt $X$ och $Y$ vara två stokastiska variabler sådana att $E(X) = \mu_X$, $V(X) = \sigma_X^2$, $E(Y) = \mu_Y$ samt $V(Y) = \sigma_Y^2$. Kovariansen $C(X, Y)$ definieras enligt

$$C(X, Y) = E\big((X - \mu_X)(Y - \mu_Y)\big)$$

och korrelationen mellan $X$ och $Y$ enligt

$$\rho(X, Y) = \frac{C(X, Y)}{\sigma_X\, \sigma_Y}.$$

Både kovarians och korrelation är ett mått på linjärt beroende mellan $X$ och $Y$ där korrelationen är normerad så det går att jämföra olika fall.

### Okorrelerad

**Definition.** Om $C(X, Y) = 0$ kallas $X$ och $Y$ för okorrelerade.

Kovariansen uppfyller följande egenskaper.

(i) $C(X, Y) = E(XY) - E(X)E(Y)$.

(ii) Om $X$ och $Y$ är oberoende så är $C(X, Y) = 0$.

(iii) $|\rho(X, Y)| \leq 1$ med likhet om och endast om det finns ett linjärt samband mellan $X$ och $Y$.

(iv) $C(X, X) = V(X)$.

(v) $C\left(a_0 + \sum_{i=1}^{m} a_i X_i,\, b_0 + \sum_{j=1}^{n} b_j Y_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j C(X_i, Y_j)$.

**Bevis**.

(i) Detta följer från i princip samma argument som Steiners sats. Låt oss expandera definition av kovarians:

$$C(X, Y) = E\big((X - \mu_X)(Y - \mu_Y)\big) = E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y = E(XY) - \mu_X \mu_Y,$$

vilket är precis vad vi ville visa.

(ii) Eftersom $X$ och $Y$ är oberoende, så följer det att $E(XY) = E(X)E(Y)$. Detta leder givetvis med föregående punkt i tanken till att $C(X, Y) = 0$.

(iii) Väntevärdet av en kvadrat är givetvis icke-negativt, så

$$
\begin{aligned}
0 &\leq E\left(\left(\frac{X-\mu_X}{\sigma_X} \pm \frac{Y-\mu_Y}{\sigma_Y}\right)^2\right) \\
&= \frac{1}{\sigma_X^2}E((X-\mu_X)^2) + \frac{1}{\sigma_Y^2}E((Y-\mu_Y)^2) \pm 2\frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X\sigma_Y} \\
&= 2\left(1 \pm \rho(X,Y)\right).
\end{aligned}
$$

Alltså måste $|\rho(X,Y)| \leq 1$. Dessutom ser vi på köpet att om $\rho(X,Y) = \pm 1$ så måste väntevärdet av kvadraten vara lika med noll, så

$$
\frac{X-\mu_X}{\sigma_X} \pm \frac{Y-\mu_Y}{\sigma_Y} = 0 \quad \Leftrightarrow \quad Y = \mu_Y \pm \frac{\sigma_Y(X-\mu_X)}{\sigma_X} \quad \Leftrightarrow \quad Y = aX + b.
$$

Således är beroendet mellan $X$ och $Y$ i form av en rät linje. En lite kvalifikation är nödvändig: det är en integral som blir noll och vi drar slutsatsen att integranden är noll. Detta är inte helt självklart. Nu råkar vi veta att integranden är icke-negativ, så vi har ingen negativ area som spökar. Men vi kan fortfarande modifiera integranden här och där utan att ändra integralen, så viss försiktighet är nödvändig. Formellt heter detta att likheten gäller *nästan överallt*. Men vi är tillbaka till Lebesgue-integralen nu, så låt oss lämna området kvickt.

(iv) Vi ser direkt att

$$
C(X,X) = E(X^2) - E(X)^2 = V(X)
$$

vilket är variansen enligt Steiners sats.

(v) Ohyggligt åbäke som lämnas som övning.. skämt åsido så kommer ni stöta på detta igen, men då med metoder från linjär algebra i bagaget så saker och ting faktiskt går att hantera utan tårar.

Observera att $C(X,Y) = 0$ *inte* nödvändigtvis innebär oberoende. Låt till exempel $X$ vara rektangelfördelad enligt $X \sim \mathrm{Re}(-1,1)$ och definiera $Y = X^2$. Uppenbarligen beroende variabler, men

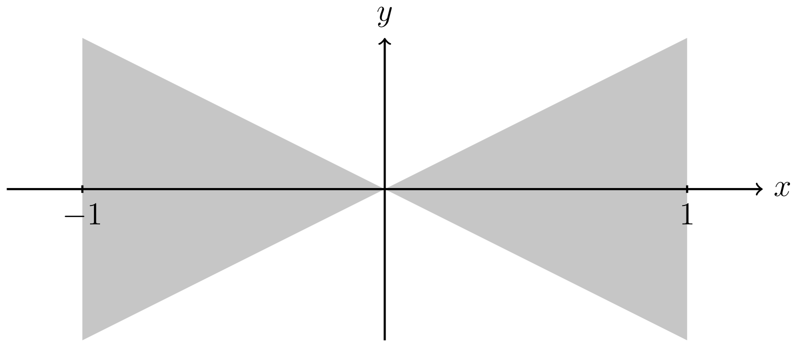
$$
C(X,Y) = E(XY) - E(X)E(Y) = E(X^3) - 0 \cdot E(Y) = E(X^3) = \int_{-1}^{1} x^3 \cdot \frac{1}{2}\,dx = 0,
$$

så $X$ och $Y$ är okorrelerade.

**Exempel**

Låt $(X,Y)$ var likformigt fördelad på området $-1 \leq x \leq 1$ och $-|x| \leq y \leq |x|$. Beräkna $E(X)$, $E(Y)$, $C(X,Y)$ och $f_X(x)$. Går det att hitta $f_{Y|X=x}(y\,;\,x)$?

**Lösning**. Låt $D$ vara området $-1 \leq x \leq 1$, $-|x| \leq y \leq |x|$. Området ges av figuren nedan. Den totala arean blir 2, så $f(x, y) = 1/2$ för $(x, y) \in D$ och $f(x, y) = 0$ utanför.




Av symmetriskäl så kommer $E(X) = E(Y) = 0$. För att hitta $C(X, Y)$ behöver vi beräkna $E(XY)$:

$$E(XY) = \int_{-1}^{0}\int_{x}^{-x} \frac{xy}{2}\,dydx + \int_{0}^{1}\int_{-x}^{x} \frac{xy}{2}\,dydx = 0,$$

även här på grund av symmetriskäl. Alltså kommer $C(X, Y) = 0 - 0 \cdot 0 = 0$. Den marginella täthethet $f_X(x)$ hittar vi genom

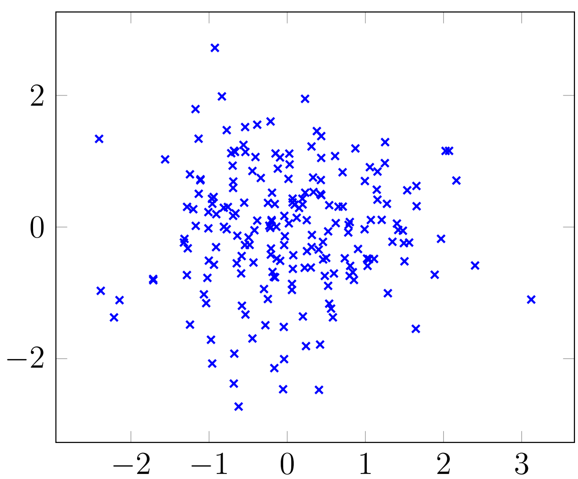
$$f_X(x) = \int_{-|x|}^{|x|} \frac{1}{2}\,dy = \frac{2|x|}{2} = |x|, \quad -1 \leq x \leq 1.$$

Den betingade fördelningen har då utseendet $f_{Y\,|\,X=x}(y) = \dfrac{1}{2|x|}$ för $-|x| \leq y \leq |x|$, åtminstone så länge $x \neq 0$. Vad händer då när $x = 0$? Vi skulle i så fall dela med något som är noll, vilket givetvis inte är tillåtet. Å andra sidan är det i detta fall nödvändigt att $y = 0$ också. En problematisk punkt som vi gör bäst i att undvika. Kom igåg att vi krävde att den marginella tätheten var strikt större än noll när vi definierade den betingade fördelningen.

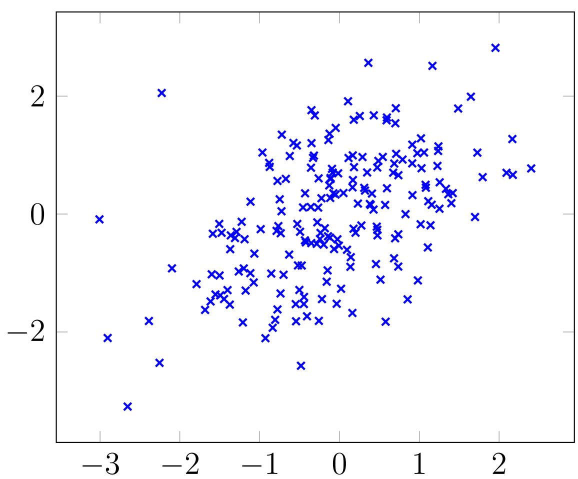
## 7.3 Vad innebär korrelationen grafiskt?

Korrelation används ofta för att avgöra om det finns ett linjärt samband mellan $x$ och $y$-värden när vi på något sätt fått en samling data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. Även detta återkommer i nästa kurs!

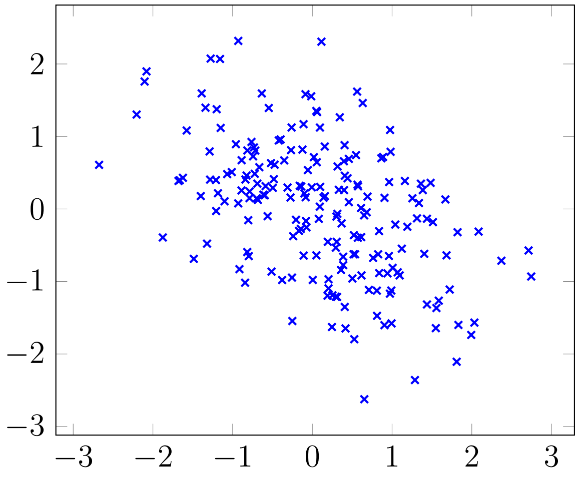
$\boldsymbol{\rho \approx 0.01}$



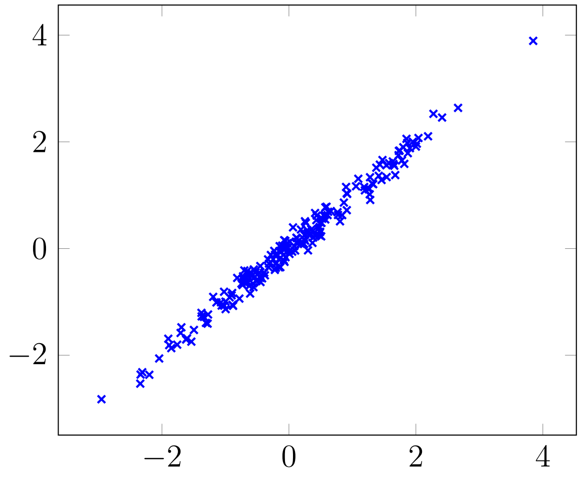

$\boldsymbol{\rho \approx 0.5}$



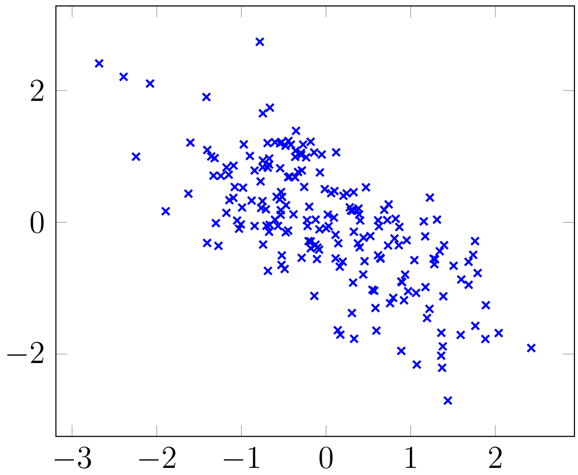
$\boldsymbol{\rho \approx -0.5}$



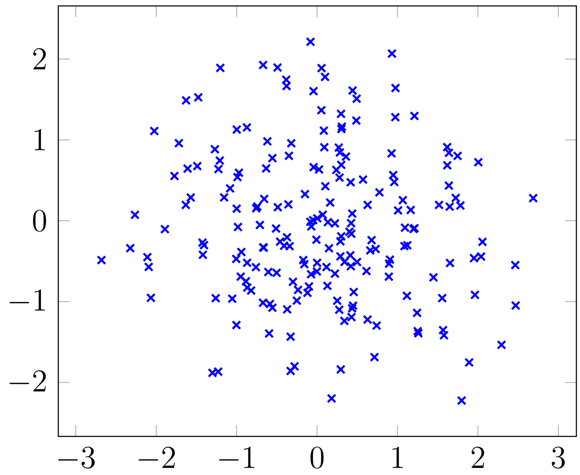
$\boldsymbol{\rho \approx 0.99}$



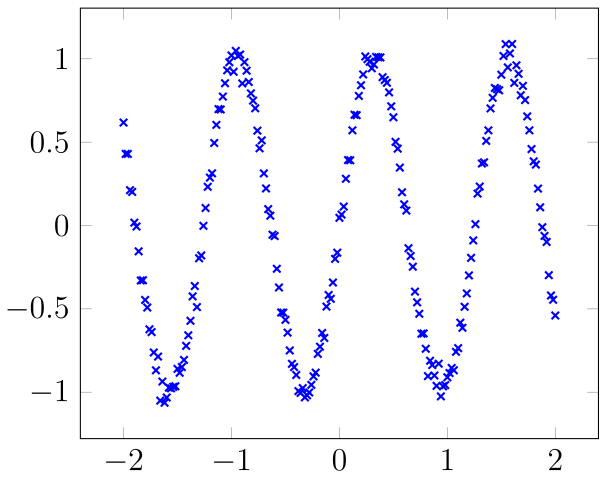
$\boldsymbol{\rho \approx -0.7}$



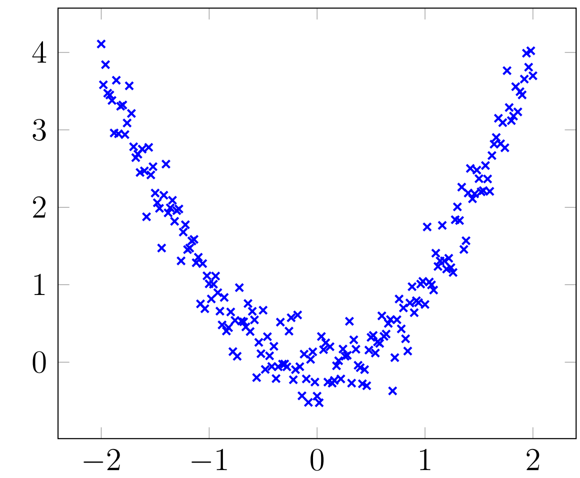
$\boldsymbol{\rho \approx -0.1}$



$\boldsymbol{\rho \approx 0.2}$ (wtf?)



$\boldsymbol{\rho \approx 0}$



## 7.4 Varians och kovarians?

När vi betraktar $n$-dimensionella stokastiska variabler så använder man ofta *kovariansmatrisen*.

**Kovariansmatris**

**Definition.** Om $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ är en $n$-dimensionell stokastisk variabel definierar vi **kovariansmatrisen $\boldsymbol{\Sigma}$** för $\boldsymbol{X}$ enligt

$$\boldsymbol{\Sigma} = \begin{pmatrix} C(X_1, X_1) & C(X_1, X_2) & \cdots & C(X_1, X_n) \\ C(X_2, X_1) & C(X_2, X_2) & \cdots & C(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ C(X_n, X_1) & C(X_n, X_2) & \cdots & C(X_n, X_n) \end{pmatrix}.$$

Notera att diagonalen på $\boldsymbol{\Sigma}$ består av varianserna. Det är också tydligt att $\Sigma$ är en diagonalmatris om komponenterna $X_1, X_2, \ldots, X_n$ är okorrelerade. Detta blir speciellt intressant för normalfördelning, men det är något som återkommer i nästa kurs.

## 7.5 Bivariat normalfördelning

Specialfallet när $n = 2$ för normalfördelningen förtjänar ett par kommentarer eftersom den situationen frekvent dyker upp. Flerdimensionell normalfördelning kommer vara centralt i nästa kurs och ni kommer få gå igenom saker ordentligt där. Vi säger att $(X, Y)$ är normalfördelad med väntevärdesvektor $\boldsymbol{\mu}$ och kovariansmatris $\boldsymbol{\Sigma}$ enligt

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \quad \text{och} \quad \begin{pmatrix} \sigma_X^2 & C(X,Y) \\ C(Y,X) & \sigma_Y^2 \end{pmatrix} = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}.$$

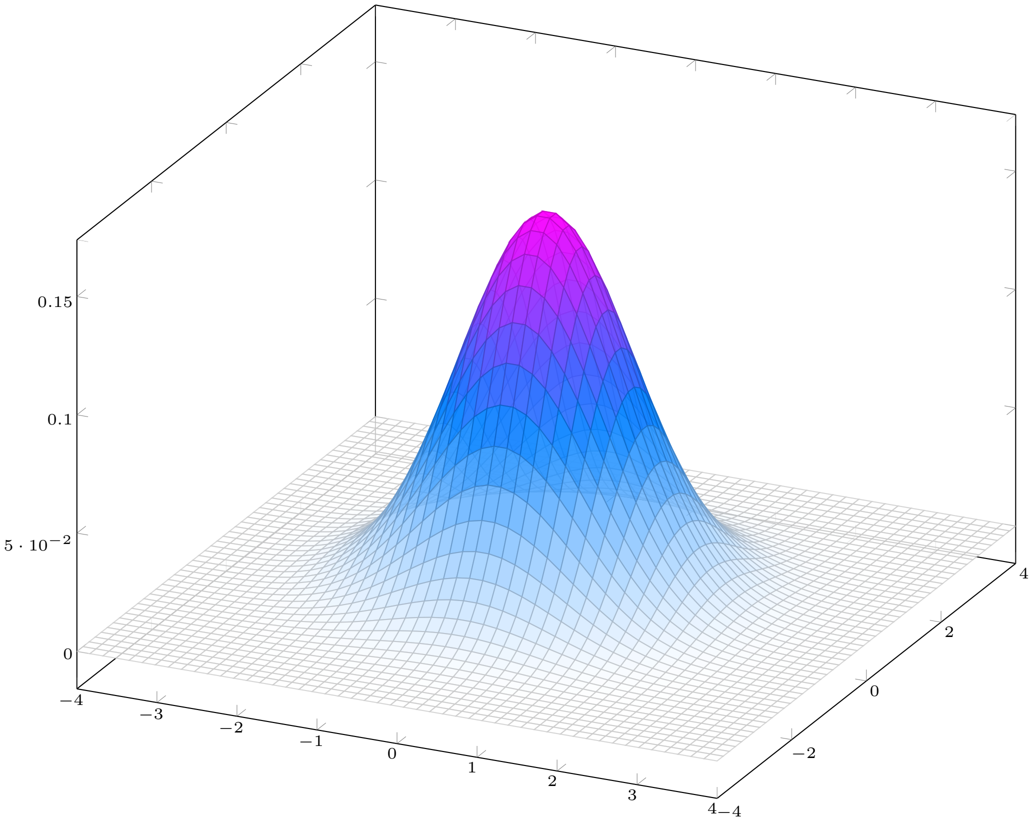
Täthetsfunktionen ges då av

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{x-\mu_X}{\sigma_X}\frac{y-\mu_Y}{\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)\right),$$

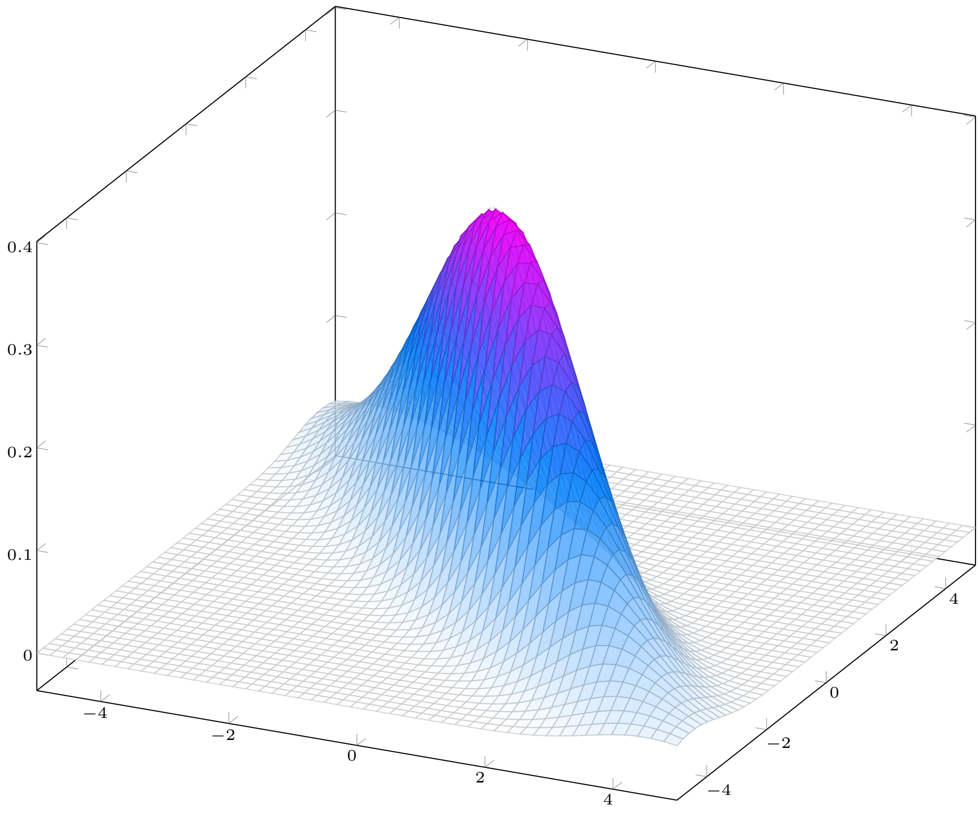
för $(x,y) \in \mathbf{R}^2$, om $|\rho| \neq 1$ (annars blir fördelningen degenererad).

Vi ser direkt att om $\rho = 0$ blir det produkten av täthetsfunktionerna för två oberoende variabler, precis som satsen i föregående avsnitt påstod. Men vad händer om variablerna inte är oberoende, dvs om $\rho \neq 0$ (oberoende och okorrelerade är ekvivalent i *normalfördelningsfallet*)?

Hur ser bivariata normalfördelningar ut? Om $\sigma_X = \sigma_Y = 1$ och $\rho = 0$ får vi följande figur:




och med $\sigma_X = \sigma_Y = 1$ och $\rho = 0.9$ erhåller vi




Låt oss beräkna den marginella tätheten $f_X(x)$. För att underlätta notationen låter vi

$$u = \frac{x - \mu_X}{\sigma_X} \quad \text{och} \quad v = \frac{y - \mu_Y}{\sigma_Y}.$$

Vi har nu

$$\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{x-\mu_X}{\sigma_X}\,\frac{y-\mu_Y}{\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 = u^2 - 2\rho uv + v^2 = (v-\rho u)^2 + (1-\rho^2)u^2,$$

så

$$\begin{aligned}
f_X(x) &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2}u^2\right)\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2(1-\rho^2)}(v-\rho u)^2\right)\,dy \\
&= \frac{1}{\sqrt{2\pi}\sigma_X}\exp\left(-\frac{1}{2}u^2\right)\frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2(1-\rho^2)}(v-\rho u)^2\right)\,dv \\
&= \frac{1}{\sqrt{2\pi}\sigma_X}\exp\left(-\frac{1}{2}u^2\right),
\end{aligned}$$

ty

$$\frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2(1-\rho^2)}(v-\rho u)^2\right)\,dv = 1.$$

### 7.5.1 Teoretiska regressionlinjer

Låt $(X, Y)$ vara bivariat normalfördelad. Då har $(X, Y)$ en simultan täthetsfunktion $f(x, y)$ och den betingade (på $X = x$) täthetsfunktionen blir

$$f_{Y\,|\,X=x}(y\,;\,x) = \frac{f(x,y)}{f_X(x)} = \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\,(v-\rho u)^2\right),$$

vilket är tätheten för en normalfördelad variabel $Y \,|\, X = x$ med

$$E(Y \,|\, X = x) = \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}\mu_X + \rho\frac{\sigma_Y}{\sigma_X}x = \beta_0 + \beta_1 x$$

och

$$V(Y \,|\, X = x) = \sigma_Y^2(1-\rho^2).$$

Det betingade (för givet $X$) väntevärdet är alltså en rät linje $y = \beta_0 + \beta_1 x$. Den observante läsaren funderar nog även om detta har med regressionsanalysen att göra, vilket ni komer att se i nästa kurs. Notera även specifikt att

$$\beta_1 = \rho\,\frac{\sigma_Y}{\sigma_X},$$

ett samband som är användbart om man vill relatera $\rho$ till välstuderade regressionskoefficienter.