1. Sannsynlighet

La $A, B, A_1, A_2, \ldots, B_1, B_2, \ldots$ være begivenheter, dvs. delmengder av et utfallsrom S .

a) Aksiomene:

$$
\begin{aligned}
&P(\mathrm{S}) = 1 \\
&P(A) \geq 0 \\
&P(A_1 \cup A_2) = P(A_1) + P(A_2) \quad \text{hvis } A_1 \cap A_2 = \emptyset \\
&P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i) \quad \text{hvis } A_i \cap A_j = \emptyset \text{ for } i \neq j
\end{aligned}
$$

b) $P(A^c) = 1 - P(A)$

c) $P(\emptyset) = 0$

d) $A \subset B \Rightarrow P(A) \leq P(B)$

e) Addisjonsetningen:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

f) Betinget sannsynlighet:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{hvis } P(B) > 0$$

g) Total sannsynlighet:

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i) \quad \text{hvis } \bigcup_{i=1}^{n} B_i = \mathrm{S} \text{ og } B_i \cap B_j = \emptyset \text{ for } i \neq j$$

h) Bayes' setning:

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)} \quad \text{under samme betingelser som i g)}$$

i) A og B er uavhengige begivenheter hvis $P(A \cap B) = P(A)P(B)$

j) $A_1, \ldots, A_n$ er uavhengige begivenheter dersom

$$P(A_{i_1} \cap \cdots \cap A_{i_m}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_m})$$

for alle delmengder av indekser $i_1, i_2, \ldots, i_m$

2. Kombinatorikk

a) To operasjoner som kan gjøres på henholdsvis n og m måter kan kombineres på $n \cdot m$ måter.

b) Antall ordnete utvalg med tilbakelegging av r elementer fra en mengde med n elementer er n^r

c) Antall ordnete utvalg uten tilbakelegging av r elementer fra en mengde med n elementer er $n(n-1)\cdots(n-r+1)$

d) Antall måter n elementer kan ordnes i rekkefølge på (permuteres) er $n! = 1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n$

e) Antall ikke-ordete utvalg av r elementer fra en mengde med n elementer er

$$\binom{n}{r} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{r!\,(n-r)!}$$

3. Sannsynlighetsfordelinger

a) For en stokastisk variabel X (diskret eller kontinuerlig) er den kumulative fordelingsfunksjonen $F(x) = P(X \leq x)$

b) For en diskret stokastisk variabel X som kan anta verdiene $x_1, x_2, x_3, \ldots$ har vi

$$p(x_j) = P(X = x_j)$$
$$F(x) = \sum_{x_j \leq x} p(x_j)$$

Betingelsene for at $p(x_j)$ skal være en punktsannsynlighet er

$$p(x_j) \geq 0 \quad \text{for alle } j$$
$$\sum_j p(x_j) = 1$$

c) For en kontinuerlig stokastisk variabel X har vi

$$P(a < X < b) = \int_a^b f(x)dx$$
$$F(x) = \int_{-\infty}^x f(u)du$$
$$f(x) = F'(x)$$

Betingelsene for at $f(x)$ skal være en sannsynlighetstetthet er

$$f(x) \geq 0$$
$$\int_{-\infty}^{\infty} f(x)dx = 1$$

d) For to stokastiske variabler X og Y (diskrete eller kontinuerlige) er den simultane kumulative fordelingsfunksjonen $F(x, y) = P(X \leq x, Y \leq y)$

e) For diskrete stokastiske variabler X og Y som kan anta henholdsvis verdiene $x_1, x_2, \ldots$ og $y_1, y_2, \ldots$ har vi

$$p(x_i, y_j) = P(X = x_i, Y = y_j)$$
$$F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p(x_i, y_j)$$

Betingelsene for at $p(x_i, y_j)$ skal være en simultan punktsannsynlighet er analoge til betingelsene i b)

f) For kontinuerlige stokastiske variabler X og Y har vi

$$P\,(\,(X,Y)\in A\,) = \int\int_A f(u,v)dvdu$$

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dvdu$$

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

Betingelsene for at $f(x,y)$ skal være en simultan sannsynlighetstetthet er analoge til betingelsene i c)

g) Marginale punktsannsynligheter:

$$p_X(x_i) = \sum_j p(x_i,y_j) \qquad \text{(for } X\text{)}$$

$$p_Y(y_j) = \sum_i p(x_i,y_j) \qquad \text{(for } Y\text{)}$$

h) Marginale sannsynlighetstettheter:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy \qquad \text{(for } X\text{)}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx \qquad \text{(for } Y\text{)}$$

i) Uavhengighet:

De stokastiske variablene X og Y er uavhengige dersom

$$p(x_i,y_j) = p_X(x_i)p_Y(y_j) \qquad \text{(diskret)}$$

$$f(x,y) = f_X(x)f_Y(y) \qquad \text{(kontinuerlig)}$$

j) Betingete punktsannsynligheter:

$$p_{X|Y}(x_i|y_j) = \frac{p(x_i,y_j)}{p_Y(y_j)} \qquad \text{(for } X \text{ gitt } Y = y_j\text{)}$$

$$p_{Y|X}(y_j|x_i) = \frac{p(x_i,y_j)}{p_X(x_i)} \qquad \text{(for } Y \text{ gitt } X = x_i\text{)}$$

Det forutsettes at $p_Y(y_j) > 0$ og $p_X(x_i) > 0$, henholdsvis. De betingete punktsannsynlighetene kan behandles som vanlige punktsannsynligheter.

k) Betingete sannsynlighetstettheter:

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} \qquad \text{(for } X \text{ gitt } Y = y\text{)}$$
$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} \qquad \text{(for } Y \text{ gitt } X = x\text{)}$$

Det forutsettes at $f_Y(y) > 0$ og $f_X(x) > 0$, henholdsvis. De betingete sannsynlighetstetthetene kan behandles som vanlige synlighetstettheter.

4. Forventning

a) Forventningsverdien til en stokastisk variabel X er definert ved

$$\mathrm{E}(X) = \sum_j x_j p(x_j) \qquad \text{(diskret)}$$
$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x) dx \qquad \text{(kontinuerlig)}$$

b) For en reell funksjon $g(X)$ av en stokastisk variabel X er

$$\mathrm{E}[g(X)] = \sum_j g(x_j) p(x_j) \qquad \text{(diskret)}$$
$$\mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx \qquad \text{(kontinuerlig)}$$

c) $\mathrm{E}(a + bX) = a + b\mathrm{E}(X)$

d) For en reell funksjon $g(X, Y)$ av to stokastiske variabler X og Y er

$$\mathrm{E}\big[g(X,Y)\big] = \sum_i \sum_j g(x_i, y_j) f(x_i, y_j) \qquad \text{(diskret)}$$
$$\mathrm{E}\big[g(X,Y)\big] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y) dy dx \qquad \text{(kontinuerlig)}$$

e) Hvis X og Y er uavhengige er $\mathrm{E}\big[g(X)h(Y)\big] = \mathrm{E}\big[g(X)\big] \cdot \mathrm{E}\big[h(Y)\big]$

f) Hvis X og Y er uavhengige er $\mathrm{E}(XY) = \mathrm{E}(X) \cdot \mathrm{E}(Y)$

g) $\mathrm{E}\left(a + \sum_{i=1}^{n} b_i X_i\right) = a + \sum_{i=1}^{n} b_i \mathrm{E}(X_i)$

h) Betinget forventning:

$$\mathrm{E}(Y|X = x_i) = \sum_j y_j p_{Y|X}(y_j|x_i) \qquad \text{(diskret)}$$

$$\mathrm{E}(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy \qquad \text{(kontinuerlig)}$$

5. Varians og standardavvik

a) Variansen og standardavviket til en stokastisk variabel X er definert ved

$$\mathrm{Var}(X) = \mathrm{E}[(X - \mu)^2]$$
$$\mathrm{sd}(X) = \sqrt{\mathrm{Var}(X)}$$

b) $\mathrm{Var}(X) = \mathrm{E}(X^2) - \big(\mathrm{E}(X)\big)^2$

c) $\mathrm{Var}(a + bX) = b^2\,\mathrm{Var}(X)$

d) Hvis $X_1, \ldots, X_n$ er uavhengige har vi

$$\mathrm{Var}\left(a + \sum_{i=1}^{n} b_i X_i\right) = \sum_{i=1}^{n} {b_i}^2\,\mathrm{Var}(X_i)$$

e)

$$\mathrm{Var}\left(a + \sum_{i=1}^{n} b_i X_i\right) = \sum_{i=1}^{n} {b_i}^2\,\mathrm{Var}(X_i) + \sum_{i=1}^{n} \sum_{j \neq i} b_i b_j \mathrm{Cov}(X_i, X_j)$$

6. Kovarians og korrelasjon

a) La X og Y være stokastiske variabler med $\mu_X = \mathrm{E}(X)$, $\sigma_X^2 = \mathrm{Var}(X)$, $\mu_Y = \mathrm{E}(Y)$ og $\sigma_Y^2 = \mathrm{Var}(Y)$. Da er kovariansen og korrelasjonen til X og Y definert ved

$$\mathrm{Cov}(X, Y) = \mathrm{E}\big[(X - \mu_X)(Y - \mu_Y)\big]$$
$$\rho = \mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

b) $\text{Cov}(X, X) = \text{Var}(X)$

c) $\text{Cov}(X, Y) = \text{E}(XY) - \text{E}(X)\text{E}(Y)$

d) X, Y uavhengige $\Rightarrow \text{Cov}(X, Y) = 0$

e)
$$\text{Cov}\left(a + \sum_{i=1}^{n} b_i X_i\,,\; c + \sum_{j=1}^{m} d_j Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} b_i d_j \text{Cov}(X_i, Y_j)$$

f) $-1 \leq \text{Corr}(X, Y) \leq 1$ og $\text{Corr}(X, Y) = \pm 1$ hvis og bare hvis det finnes to tall a, b slik at $Y = a + bX$ (bortsett, eventuelt, på et område med sannsynlighet 0)

8. Noen diskrete sannsynlighetsfordelinger

a) Binomisk fordeling:

Punktsannsynlighet: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, \ldots, n$

Forventning: $\text{E}(X) = np$

Varians : $\text{Var}(X) = np(1-p)$

Tilnærmelse 1: $Z = \dfrac{X - np}{\sqrt{np(1-p)}}$ er tilnærmet standard normalfordelt når np og $n(1-p)$ begge er tilstrekkelig store (minst 10)

Tilnærmelse 2: X er tilnærmet Poisson fordelt med parameter $\lambda = np$ når n er stor og p er liten

Addisjonsregel: $X \sim$ binomisk (n, p), $Y \sim$binomisk (m, p)
og X, Y uavhengige $\Rightarrow X + Y \sim$ binomisk $(n + m, p)$

b) Geometrisk fordeling:

Punktsannsynlighet: $P(X = k) = (1 - p)^{k-1}p \qquad k = 1, 2, \ldots$

Forventning: $\mathrm{E}(X) = 1/p$

Varians: $\mathrm{Var}(x) = (1 - p)/p^2$

Addisjonsregel: Summen av r uavhengige geometriske
fordelte variabler er negativt binomisk fordelt

d) Hypergeometrisk fordeling:

Punktsannsynlighet: $P(X = k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$

Forventning: $\mathrm{E}(X) = m \cdot \frac{r}{n}$

Varians: $\mathrm{Var}(X) = m\frac{r}{n}(1 - \frac{r}{n})\frac{n-m}{n-1}$

Tilnærmelse: X er tilnærmet binomisk $(m, \frac{r}{n})$ når m er mye mindre enn n

e) Poisson fordelingen:

Punktsannsynlighet: $P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda} \qquad k = 0, 1, \ldots$

Forventning: $\mathrm{E}(X) = \lambda$

Varians: $\mathrm{Var}(X) = \lambda$

Tilnærmelse: $Z = \dfrac{X - \lambda}{\sqrt{\lambda}}$ er tilnærmet standard normalfordelt
når λ er tilstrekkelig stor (minst 10)

Addisjonsregel: $X \sim$Poisson (λ_1), $Y \sim$Poisson (λ_2)
og X, Y uavhengige $\Rightarrow X + Y \sim$ Poisson $(\lambda_1 + \lambda_2)$

9. Noen kontinuerlige sannsynlighetsfordelinger

a) Normalfordelingen:

Tetthet: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \qquad -\infty < x < \infty$

Forventning: $\mathrm{E}(X) = \mu$

Varians: $\mathrm{Var}(X) = \sigma^2$

Transformasjon: $X \sim N(\mu, \sigma^2) \Rightarrow a + bX \sim N(a + b\mu, b^2\sigma^2)$
$X \sim N(\mu, \sigma^2) \Rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$

Addisjonsregel: $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, X, Y uavhengige
$\Rightarrow X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

b) Eksponentialfordelingen:

Tetthet: $f(x) = \lambda e^{-\lambda x} \qquad x > 0$

Forventning: $\mathrm{E}(X) = 1/\lambda$

Varians: $\mathrm{Var}(X) = 1/\lambda^2$

c) Gammafordelingen:

Tetthet: $f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \qquad x > 0$

Gammafunksjonen: $\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du$
$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$
$\Gamma(n) = (n-1)!$ når n er et helt tall
$\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$

Forventning: $\mathrm{E}(X) = \alpha\beta$

Varians: $\mathrm{Var}(X) = \alpha\beta^2$

d) Kji-kvadratfordelingen:

Tetthet: $f(v) = \frac{1}{2^{n/2}\Gamma(n/2)} v^{(n/2)-1} e^{-v/2} \qquad v > 0$

n er antall frihetsgrader

Forventning: $\mathrm{E}(V) = n$

Varians: $\mathrm{Var}(V) = 2n$

e) Students t-fordeling:

Tetthet: $f(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\Gamma(n/2)} (1 + \frac{t^2}{n})^{-(n+1)/2} \qquad -\infty < t < \infty$

n er antall frihetsgrader

Forventning: $\mathrm{E}(T) = 0 \qquad (n \geq 2)$

Varians: $\mathrm{Var}(T) = n/(n-2) \qquad (n \geq 3)$

Resultat: $Z \sim N(0,1), \quad U \sim \chi^2_n, \quad Z, U$ uavhengige $\Rightarrow Z/\sqrt{U/n} \sim t_n$

10. Ett normalfordelt utvalg

Hvis $X_1, X_2, \ldots, X_n$ er uavhengige og $N(\mu, \sigma^2)$-fordelte så har vi at:

a) $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ og $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$ er uavhengige

b) $\overline{X} \sim N(\mu, \sigma^2/n)$

d) $\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

Konfidensintervall og testing

Binomisk

$$\hat{p} = \frac{X}{n}$$

Konfidensintervall

$$\left[\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

Testing

$$H_0 : p = p_0 \qquad H_1 : \begin{array}{l} p < p_0 \\ p \neq p_0 \\ p > p_0 \end{array} \qquad Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \qquad \text{Forkast} \quad \begin{array}{l} Z \leq -z_\alpha \\ Z \leq -z_{\frac{\alpha}{2}}, Z \geq z_{\frac{\alpha}{2}} \\ Z \geq z_\alpha \end{array}$$

I tosidig situasjon: kan bruke konfidensintervall i stedet. Husk definisjon og bruk av p-verdi her og ellers

Målemodell $X_1, \ldots, X_n$ uavhengige og normalfordelte

$$E(X_i) = \mu \qquad \mathrm{Var}(X_i) = \sigma^2$$

Konfidensintervall

$$\hat{\mu} = \overline{X}$$

σ kjent $\qquad \left[\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$

σ ukjent $\qquad \left[\overline{X} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \overline{X} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right]$

$(n-1)$ frihetsgrader

Testing

σ **kjent**

$$H_0 : \mu = \mu_0 \qquad H_1 : \begin{array}{l} \mu < \mu_0 \\ \mu \neq \mu_0 \\ \mu > \mu_0 \end{array} \qquad Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \qquad \text{Forkast} \; \begin{array}{l} Z \leq -z_\alpha \\ Z \leq -z_{\frac{\alpha}{2}}, Z \geq z_{\frac{\alpha}{2}} \\ Z \geq z_\alpha \end{array}$$

σ **ukjent**

$$H_0 : \mu = \mu_0 \qquad H_1 : \begin{array}{l} \mu < \mu_0 \\ \mu \neq \mu_0 \\ \mu > \mu_0 \end{array} \qquad t = \frac{\overline{X} - \mu_0}{\frac{s}{\sqrt{n}}} \qquad \text{Forkast} \; \begin{array}{l} t \leq -t_\alpha \\ t \leq -t_{\frac{\alpha}{2}}, t \geq t_{\frac{\alpha}{2}} \\ t \geq t_\alpha \end{array}$$