

## 92MA31/37: Formelblad

Johan Thim\*

11 september 2013

### Kombinatorik

Om  $n$  är ett icke-negativt heltal definieras  $n!$  som  $n \cdot (n-1) \cdot (n-2) \cdots 1$  om  $n \geq 1$  och  $0! = 1$ . Talet  $n!$  beskriver på hur många sätt  $n$  distinkta objekt kan ordnas (antalet permutationer).

Låt  $k$  och  $n$  vara icke-negativa heltal så att  $k \leq n$ . Binomialkoefficienten  $n$  över  $k$  definieras som

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}.$$

Att välja  $k$  element av  $n$  kan göras på följande antal sätt:

	Med återläggning	Utan återläggning
Med ordning	$n^k$	$\frac{n!}{(n-k)!}$
Utan ordning	$\frac{(n+k-1)!}{k!(n-1)!}$	$\frac{n!}{(n-k)!k!}$

### Sannolikhetslära

Låt  $A$  och  $B$  vara händelser i ett utfallsrum  $\Omega$  med sannolikhetsmått  $P$ . Följande samband gäller:

1. Komplementhändelse:

$$P(A^*) = 1 - P(A).$$

2. Unionen:

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

3. Händelserna  $A$  och  $B$  är oberoende om och endast om

$$P(A \cap B) = P(A)P(B).$$

4. Om  $P(A) > 0$  ges den betingade sannolikheten för  $B$  (betingat på att  $A$  inträffar) av

$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

5. Lagen om total sannolikhet (LOTS): Låt  $H_1, H_2, \dots, H_n$  vara parvis disjunkta händelser ( $H_i \cap H_j = \emptyset$  om  $i \neq j$ ) så att  $H_1 \cup H_2 \cup \dots \cup H_n = \Omega$ . Då gäller att

$$P(A) = \sum_{i=1}^n P(A|H_i)P(H_i).$$

6. Bayes sats: med samma förutsättningar som för LOTS gäller att

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{\sum_{k=1}^n P(A|H_k)P(H_k)}.$$

\*jothi@mai.liu.se

### Väntevärde och varians

För en slumpvariabel  $X$  och en funktion  $g(x)$  gäller att

$$E(g(X)) = \sum_k g(k)p_X(k)$$

om  $X$  är diskret och

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f_X(x) dx$$

om  $X$  är kontinuerlig. Variansen fås genom

$$V(X) = E((X - E(X))^2) = E(X^2) - E(X)^2.$$

Den sista likheten brukar kallas Steiners sats.

Om  $X_1, X_2, \dots, X_n$  är stokastiska variabler och  $a_1, a_2, \dots, a_n, b$  är reella konstanter så gäller att

$$E\left(b + \sum_{k=1}^n a_k X_k\right) = b + \sum_{k=1}^n a_k E(X_k).$$

Om variablerna dessutom är oberoende så gäller att

$$V\left(b + \sum_{k=1}^n a_k X_k\right) = \sum_{k=1}^n a_k^2 V(X_k).$$

Kovariansen  $C(X, Y)$  mellan två stokastiska variabler kan beräknas enligt

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

### Vanliga fördelningar

Låt  $X$  vara en stokastisk variabel och  $P$  sannolikhetsmåttet. Fördelningsfunktionen  $F(x)$  definieras som

$$F(x) = P(X \leq x) \text{ för alla } x \in \mathbb{R}.$$

#### Diskreta fördelningar

Sannolikhetsfunktion:  $p_X(k) = P(X = k)$ .

- Binomialfördelning:  $X \sim \text{Bin}(n, p)$ ,  $n \geq 0$  heltal,  $0 < p < 1$ ,

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \dots, n,$$

$$E(X) = np, \quad V(X) = np(1-p).$$

- Hypergeometrisk:  $X \sim \text{Hyp}(N, n, p)$ ,  $0 < p < 1$ ,

$$p_X(k) = \frac{\binom{Np}{k} \binom{N(1-p)}{n-k}}{\binom{N}{n}}, \quad 0 \leq k \leq Np, \quad 0 < n-k \leq N(1-p),$$

$$E(X) = np, \quad V(X) = \frac{N-n}{N-1} np(1-p).$$

- Ffg:  $X \sim \text{Ffg}(p)$ ,  $0 < p < 1$ ,

$$p_X(k) = p(1-p)^{k-1}, \quad k = 1, 2, 3, \dots,$$

$$E(X) = \frac{1}{p}, \quad V(X) = \frac{1-p}{p^2}.$$

- Poisson:  $X \sim \text{Po}(\mu)$ ,  $\mu > 0$ ,

$$p_X(k) = \frac{\mu^k}{k!} e^{-\mu}, \quad k = 0, 1, 2, \dots,$$

$$E(X) = \mu, \quad V(X) = \mu.$$

## Kontinuerliga fördelningar

Täthetsfunktionen  $f_X(x)$  uppfyller:

$$P(X \leq x) = \int_{-\infty}^x f(s) ds, \quad \text{för alla } x \in \mathbb{R}.$$

- Rektangelfördelning (likformig):  $X \sim R(a, b)$ , där  $a < b$ ,

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{för övrigt,} \end{cases}$$

$$E(X) = \frac{a+b}{2}, \quad V(X) = \frac{(b-a)^2}{12}.$$

- Normalfördelning:  $X \sim N(\mu, \sigma)$ ,  $\mu \in \mathbb{R}$  och  $\sigma > 0$ ,

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R},$$

$$E(X) = \mu, \quad V(X) = \sigma^2.$$

- Exponentialfördelning:  $X \sim \text{Exp}(\mu)$ ,  $\mu > 0$ ,

$$f_X(x) = \begin{cases} \frac{1}{\mu} \exp(-x/\mu), & x \geq 0, \\ 0, & \text{för övrigt,} \end{cases}$$

$$E(X) = \mu, \quad V(X) = \mu^2.$$

Ibland används intensiteten  $\lambda = 1/\mu$  som parameter istället.

## Centrala gränsvärdessatsen

Låt  $X_1, X_2, X_3, \dots$  vara en följd av oberoende och likafördelade stokastiska variabler med  $E(X_k) = \mu$  och  $V(X_k) = \sigma^2$  för alla  $k = 1, 2, 3, \dots$ . Då gäller att

$$X = \sum_{k=1}^n X_k \stackrel{\text{appr.}}{\sim} N(n\mu, \sigma\sqrt{n})$$

för stora  $n$  i den meningen att

$$\lim_{n \rightarrow \infty} P\left(a < \frac{X - n\mu}{\sigma\sqrt{n}} < b\right) = \Phi(b) - \Phi(a)$$

för alla  $a, b \in \mathbb{R}$ . För god approximation brukar  $n \geq 30$  duga.

## Approximationer

$$\text{Hyp}(N, n, p) \stackrel{\text{appr.}}{\sim} \text{Bin}(n, p) \text{ om } \frac{n}{N} \leq \frac{1}{10}.$$

$$\text{Hyp}(N, n, p) \stackrel{\text{appr.}}{\sim} N\left(np, \sqrt{\frac{N-n}{N-1}np(1-p)}\right) \text{ om } \frac{N-n}{N-1}np(1-p) \geq 10.$$

$$\text{Bin}(n, p) \stackrel{\text{appr.}}{\sim} \text{Po}(np) \text{ om } n \geq 10 \text{ och } p \leq 0.1.$$

$$\text{Bin}(n, p) \stackrel{\text{appr.}}{\sim} N(np, \sqrt{np(1-p)}) \text{ om } np(1-p) \geq 10.$$

$$\text{Po}(\mu) \stackrel{\text{appr.}}{\sim} N(\mu, \sqrt{\mu}) \text{ om } \mu \geq 15.$$

## Statistik

### Stickprov

Vi kallar  $X_1, X_2, \dots, X_n$  för ett stickprov (av storlek  $n$ ) av en stokastisk variabel  $X$  om alla  $X_k$  är oberoende av varandra och har samma fördelning som  $X$ . Ibland kallar vi stickprovet för ett stickprov av fördelningen för  $X$ . Medelvärde och stickprovsvariansen,

$$\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k \quad \text{och} \quad S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2,$$

är väntevärdesriktiga och konsistenta skattningar av väntevärdet respektive variansen för fördelningen. För att skatta standardavvikelsen brukar  $S = \sqrt{S^2}$  användas.

Om  $X \sim N(\mu, \sigma)$  så gäller att

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \quad \text{och} \quad \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

### Normalapproximation

Om  $X \sim \text{Bin}(n, p) \stackrel{\text{appr.}}{\sim} N(np, \sqrt{np(1-p)})$ , så gäller med  $\hat{p} = X/n$  att

$$\frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \stackrel{\text{appr.}}{\sim} N(0, 1).$$

Om  $X \sim \text{Po}(\mu) \stackrel{\text{appr.}}{\sim} N(\mu, \sqrt{\mu})$ , så gäller med  $\hat{\mu} = X$  att

$$\frac{\hat{\mu} - \mu}{\sqrt{\hat{\mu}}} \stackrel{\text{appr.}}{\sim} N(0, 1).$$

### TVå stickprov

Låt  $X_1, X_2, \dots, X_n$  och  $Y_1, Y_2, \dots, Y_m$  vara två av varandra oberoende stickprov från  $N(\mu_X, \sigma)$  respektive  $N(\mu_Y, \sigma)$  (samma  $\sigma$ ). Om  $S_X^2$  och  $S_Y^2$  är de respektive stickprovsvarianserna så ges den sammanvägda skattningen av  $\sigma^2$  av

$$S^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{(n-1) + (m-1)}.$$

Om  $a$  och  $b$  är reella tal så är

$$\frac{(a\bar{X} + b\bar{Y}) - (a\mu_X + b\mu_Y)}{\sigma\sqrt{\frac{a^2}{n} + \frac{b^2}{m}}} \sim N(0, 1) \quad \text{och} \quad \frac{(a\bar{X} + b\bar{Y}) - (a\mu_X + b\mu_Y)}{S\sqrt{\frac{a^2}{n} + \frac{b^2}{m}}} \sim t(n+m-2).$$

### Enkel linjär regression

Om  $y_j$  är observationer av stokastiska variabler  $Y_j = b_0 + b_1x_j + \epsilon_j$  för  $j = 1, 2, \dots, n$ , där  $\epsilon_j \sim N(0, \sigma)$  är oberoende och  $(x_j, y_j)$  är uppmätta datapunkter, så kan de värden på  $b_0$  och  $b_1$  som minimerar

$$\sum_{j=1}^n (y_j - (b_0 + b_1x_j))^2$$

beräknas enligt följande:

$$b_0 = \bar{y} - b_1\bar{x} \quad \text{och} \quad b_1 = \frac{\sum_{j=1}^n x_j y_j - n\bar{x}\bar{y}}{\sum_{j=1}^n x_j^2 - n\bar{x}^2}.$$







