

Normalfördelningen

En stokastisk variabel X sägs vara normalfördelad med parametrar m och $\sigma > 0$ om

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{(x-m)^2/2\sigma^2}$$

för alla x . Läs gärna beviset i boken att detta är en giltig täthet, dvs. integrerar sig till 1. Kodbe-teckning X är $N(m, \sigma)$.

Detta är tvivelsutan den viktigaste och mest fundamentala fördelningen. En av anledningarna är att den har en rad trevliga matematiska egenskaper, men framför allt, det viktiga resultat som går under namnet *Centrala gränsvärdessatsen* som säger att summor av (oberoende och likafördelade) stokastiska variabler är approximativt normalfördelade.

Fördelningen är symmetrisk runt m så alltså är parametern m ingenting annat än väntevärdet, $E[X] = m$. Vidare så är

$$\begin{aligned} V(X) &= E[(X-m)^2] = \int_{-\infty}^{\infty} (x-m)^2 f_X(x) dx = \int_{-\infty}^{\infty} (x-m)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{(x-m)^2/2\sigma^2} dx \\ &= \int_{-\infty}^{\infty} u^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-u^2/2\sigma^2} du = \int_{-\infty}^{\infty} u \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-u^2/2\sigma^2} \cdot u du \\ &= \underbrace{\left[u \cdot \frac{-1}{\sqrt{2\pi}\sigma} e^{-u^2/2\sigma^2} \sigma^2 \right]_{-\infty}^{\infty}}_{=0} + \sigma^2 \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-u^2/2\sigma^2} du}_{=1} = \sigma^2. \end{aligned}$$

eftersom det sista är integralen för en täthet för en normalfördelad stokastisk variabel.

Fördelningsfunktionen för en normalfördelad stokastisk variabel X

$$F_X(t) = P(X \leq t) = \int_{-\infty}^t f_X(x) dx = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}\sigma} e^{(x-m)^2/2\sigma^2} dx$$

finns inte på någon sluten form. Vi skall se hur vi kan klara oss med fallet $m = 0$, $\sigma = 1$.

Sats. Låt $X_1, \dots, X_n$ vara en sekvens av oberoende, normalfördelade stokastiska variabler, och $a_1, \dots, a_n$ och b konstanter. Då är

$$Y = a_1 X_1 + \dots + a_n X_n + b$$

normalfördelad, Y är $N(m, \sigma)$, med väntevärde

$$m = E[Y] = E[a_1 X_1 + \dots + a_n X_n + b] = a_1 E[X_1] + \dots + a_n E[X_n] + b$$

och varians

$$\sigma^2 = V(Y) = a_1^2 V(X_1) + \dots + a_n^2 V(X_n).$$

Alltså,

$$Y \text{ är } N\left(a_1 E[X_1] + \dots + a_n E[X_n] + b, \sqrt{a_1^2 V(X_1) + \dots + a_n^2 V(X_n)}\right)$$

och linjärkombinationer av oberoende normalfördelade stokastiska variabler är normalfördelade med rätt väntevärde och rätt varians.

Om X är $N(m, \sigma)$ så har

$$Z = \frac{X - m}{\sigma}$$

väntevärde $E[Z] = E\left[\frac{X-m}{\sigma}\right] = \frac{1}{\sigma}(E[X] - m) = 0$ och varians $V(Z) = \frac{1}{\sigma^2}V(X) = 1$. Satsen säger att Z är normalfördelad och

$$f_Z(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} = \varphi(x).$$

En normalfördelad s.v. med väntevärde 0 och varians (standardavvikelse) 1 sägs vara *standardnormalfördelad*. Dess täthetsfunktion betecknas av Blom med $\varphi(x)$ och dess fördelningsfunktion med $\Phi(x) = P(Z \leq x)$.

Notera att

$$F_X(x) = P(X \leq x) = P\left(\frac{X-m}{\sigma} \leq \frac{x-m}{\sigma}\right) = \Phi\left(\frac{x-m}{\sigma}\right).$$

Så för att räkna ut fördelningsfunktionens värde i en punkt för en godtycklig normalfördelning översätter vi den till motsvarande punkt för en standardnormalfördelning. Funktionen Φ finns tabulerad i formelsamlingen.

Vi skall ofta lösa ekvationer av typen: för ett givet α finn det värde z så att

1. $P(Z > z) = P(Z \geq z) = \alpha$
2. $P(Z < z) = P(Z \leq z) = \alpha$
3. $P(|Z| > z) = P(|Z| \geq z) = \alpha$

Dessa värden på z bestäms av *kvantiler* λ_α i standardnormalfördelningen definierade av

$$1 - \Phi(\lambda_\alpha) = \alpha.$$

och finns tabulerade i boken. Man kan observera att

$$\Phi(-x) = 1 - \Phi(x)$$

så $\lambda_{1-\alpha} = -\lambda_\alpha$.

Vi har att

1. $P(Z > z) = 1 - \Phi(z) = \alpha$ ger $z = \lambda_\alpha$
2. $P(Z < z) = \Phi(z) = \alpha$ ger $z = \lambda_{1-\alpha} = -\lambda_\alpha$.
3. $P(|Z| > z) = \alpha$ ger $z = \lambda_{\alpha/2}$.

Sats (Centrala gränsvärdessatsen (CGS)). Låt $X_1, X_2, \dots$ vara en sekvens av oberoende, likafördelade stokastiska variabler med väntevärde m och standardavvikelse σ . Då gäller att

$$\lim_{n \rightarrow \infty} P\left(\frac{1}{\sigma\sqrt{n}}\left(\sum_{k=1}^n X_k - nm\right) \leq x\right) = \Phi(x)$$

Notera att variabeln

$$\frac{1}{\sigma\sqrt{n}}\left(\sum_{k=1}^n X_k - nm\right)$$

har väntevärde 0 och standardavvikelse 1 för alla n .

Man använder konvergensen för att säga att

$$\frac{\sum_{k=1}^n X_k - nm}{\sigma\sqrt{n}}$$

är approximativt $N(0, 1)$, eller

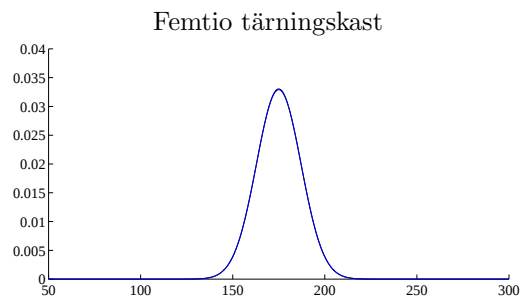
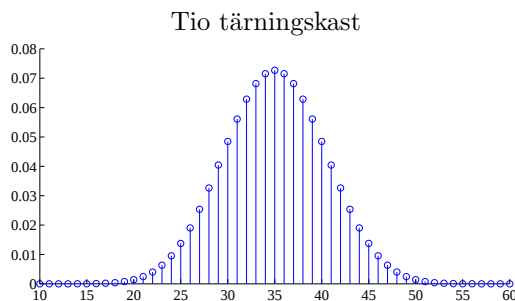
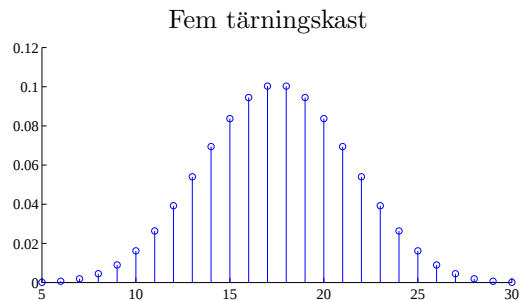
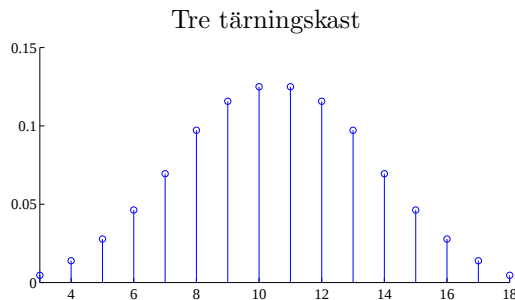
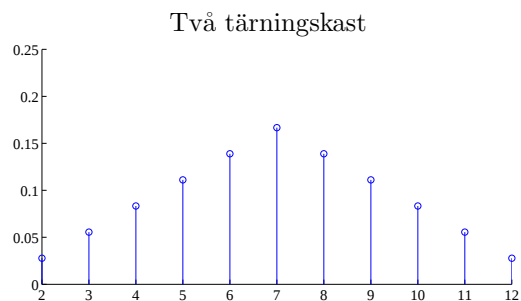
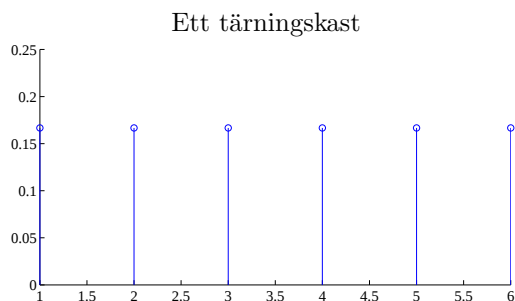
$$\sum_{k=1}^n X_i \stackrel{\text{approx}}{\text{är}} N(mn, \sqrt{n}\sigma)$$

för stora värden på n , dvs summor av stokastiska variabler är approximativt normalfördelade.

Hur stort n skall vara för att approximationen skall vara bra beror på fördelningen för de stokastiska variablerna. Symmetriska fördelningar konvergerar snabbare än assymetriska.

Anmärkning. Det är möjligt att tillåta variablerna ha olika fördelning (ett tillräckligt villkor är t.ex. om $E[X_k^\gamma] < \infty$ för något $\gamma > 2$). Man kan även tillåta ett visst beroende variablerna emellan.

Exempel: Låt $X_1, X_2, \dots$ vara resultatet av oberoende tärningskast. $P(X = k) = \frac{1}{6}$ för $k = 1, \dots, 6$. Nedan visas fördelningarna för summan av 1, 2, 3, 5, 10 och 50 tärningskast.



Binomialfördelning och dess släktingar

Hypergeometrisk fördelning

En stokastisk variabel X sägs vara *hypergeometriskt fördelad* om

$$p_X(k) = \mathbf{P}(X = k) = \frac{\binom{Np}{k} \binom{N(1-p)}{n-k}}{\binom{N}{n}}$$

för $k = \max(0, n - N(1 - p)), \dots, \min(Np, n)$, där N , Np och $N(1 - p)$ är icke-negativa heltal och $0 \leq p \leq 1$.

Modellsituation: Av N enheter välj ut n på måfå utan återläggning. Om s enheter har en viss egenskap, låt X vara antalet enheter med egenskapen bland de n utvalda. Då är enligt multiplikationsprincipen

$$\mathbf{P}(X = k) = \frac{\binom{s}{k} \binom{N-s}{n-k}}{\binom{N}{n}}.$$

Med $p = s/N$, andelen enheter med egenskapen, erhålls det tidigare uttrycket.

Man kan visa att

$$\mathbf{E}[X] = np = n \frac{s}{N}$$

och

$$\mathbf{V}(X) = np(1-p) \frac{N-n}{N-1} = n \frac{s}{N} \left(1 - \frac{s}{N}\right) \frac{N-n}{N-1}.$$