

Normalfördelningen

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

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

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

Detta är tvivelsutan en av de viktigaste fördelningarna. 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 μ så alltså är parametern μ ingenting annat än väntevärdet, $E[X] = \mu$.

En normalfördelad stokastisk variabel har momentgenererande funktion

$$\begin{aligned} m_X(s) &= E[e^{Xs}] = \int_{-\infty}^{\infty} e^{xs} f_X(x) dx = \int_{-\infty}^{\infty} e^{xs} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{-1}{2\sigma^2}(-2\sigma^2 xs + x^2 - 2\mu x + \mu^2)\right\} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{-1}{2\sigma^2}([x - (\mu + \sigma^2 s)]^2 - (\mu + \sigma^2 s)^2 + \mu^2)\right\} dx \\ &= e^{\mu s + \sigma^2 s^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{-1}{2\sigma^2}[x - (\mu + \sigma^2 s)]^2\right\} dx = e^{\mu s + \sigma^2 s^2/2}. \end{aligned}$$

Förutom att denna momentgenererande funktion entydigt bestämmer fördelningen så kan man utnyttja den för att bestämma momenten:

$$E[X^k] = \frac{d^k}{ds^k} m_X(s) \Big|_{s=0} = m_X^{(k)}(0)$$

Alltså,

$$\begin{aligned} m'_X(s) &= \frac{d}{ds} m_X(s) = (\mu + \sigma^2 s) e^{\mu s + \sigma^2 s^2/2} \\ m''_X(s) &= \frac{d^2}{ds^2} m_X(s) = \sigma^2 e^{\mu s + \sigma^2 s^2/2} + (\mu + \sigma^2 s)^2 e^{\mu s + \sigma^2 s^2/2} \end{aligned}$$

så

$$E[X] = m'_X(0) = \mu \quad E[X^2] = m''_X(0) = \sigma^2 + \mu^2$$

det vill säga $V(X) = E[X^2] - (E[X])^2 = \sigma^2$.

Fördelningsfunktionen för en normalfördelad stokastisk variabel X

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

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

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 $N(\mu, \sigma)$, med väntevärde

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

och varians

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

Bevis: Y har momentgenererande funktion

$$\begin{aligned} m_Y(s) &= E[e^{sY}] = E[e^{s(a_1 X_1 + \dots + a_n X_n + b)}] = \{\text{oberoende}\} = E[e^{sa_1 X_1}] \dots E[e^{sa_n X_n}] e^{sb} \\ &= m_{X_1}(a_1 s) \dots m_{X_n}(a_n s) e^{sb} = e^{\mu_1 a_1 s + a_1^2 \sigma_1^2 s^2 / 2} \dots e^{\mu_n a_n s + a_n^2 \sigma_n^2 s^2 / 2} e^{sb} \\ &= e^{(\mu_1 a_1 + \dots + \mu_n a_n + b)s + (a_1^2 \sigma_1^2 + \dots + a_n^2 \sigma_n^2) s^2 / 2} = e^{\mu s + \sigma^2 s^2 / 2} \end{aligned}$$

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 $N(\mu, \sigma)$ så har

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

väntevärde $E[Z] = E\left[\frac{X - \mu}{\sigma}\right] = \frac{1}{\sigma}(E[X] - \mu) = 0$ 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

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

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

I formelsamlingen finns även $N(0, 1)$ vanliga kvantiler till fördelningen tabulerade.

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

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

Notera att variabeln

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

har väntevärde 0 och standardavvikelse 1 för alla n .

Man använder konvergensen för att säga att

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

är approximativt $N(0, 1)$, eller

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

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.

Binomialfördelning och dess släktingar

Genomgående modellsituation: Betrakta oberoende försök där vid varje försök det finns en sannolikhet p att en händelse A inträffar.

Låt X beteckna antalet gånger man får göra försöket tills man ser att A inträffar för första gången.

Då är X för första gången-fördelad, skrivet X är ffg(p), om

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

för $k = 1, 2, 3, \dots$. Vi bestämmer den momentgenererande funktionen till X

$$\begin{aligned} m_X(s) &= E[e^{Xs}] = \sum_k e^{ks} P(X = k) = \sum_{k=1}^{\infty} e^{ks} (1-p)^{k-1} p = pe^s \sum_{k=0}^{\infty} (e^s(1-p))^k \\ &= pe^s \frac{1}{1 - e^s(1-p)} = \frac{p}{e^{-s} - (1-p)}. \end{aligned}$$

Alltså är

$$\frac{d}{ds} m_X(s) = \frac{pe^{-s}}{(e^{-s} - (1-p))^2} \quad m'_X(0) = \frac{1}{p} = E[X]$$

och

$$\frac{d^2}{ds^2} m_X(s) = \frac{(-1)pe^{-s}}{(e^{-s} - (1-p))^2} + \frac{2pe^{-2s}}{(e^{-s} - (1-p))^3}$$

så

$$m''_X(0) = \frac{2-p}{p^2} = E[X^2].$$

Detta ger

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