Våra vanligaste fördelningar

Matematisk statistik för D3, VT02

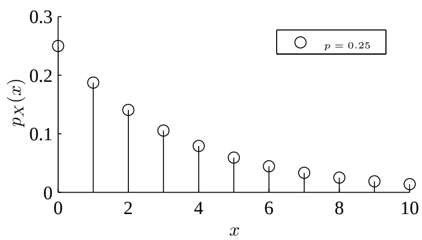
Geometrisk fördelning

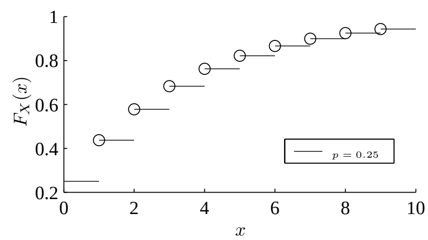
X är geometriskt fördelad med parameter p, $X \sim \mathrm{Geo}(p)$, om

$$\mathsf{P}(X = k) = (1-p)^{k-1}p \qquad\qquad \mathsf{P}(X \le k) = 1 - (1-p)^k$$

för $k = 1, 2, \ldots$ Beskriver antalet oberoende försök vi måste göra tills vi lyckas första gången, där varje försök lyckas med sannolikheten p. Minneslös på samma sätt som exponentialfördelningen.

$$\mathsf{E}[X] = \frac{1}{p} \qquad\qquad \mathrm{Var}(X) = \frac{1-p}{p^2}.$$





Exponentialfördelningen

X är exponentialfördelad med parameter $\lambda > 0$, $X \sim \exp(\lambda)$, om

$$f_X(x) = \lambda e^{-\lambda x},\ x \ge 0 \qquad\qquad F_X(x) = 1 - e^{-\lambda x},\ x \ge 0.$$

Exponentialfördelningen beskriver tiden tills en händelse inträffar, där λ är intensiteten för händelser. Den förväntade tiden mellan händelser är $1/\lambda$, och kursboken använder $\beta = 1/\lambda$ som parameter. Vanligtvis uppträder exponentialfördelningen som livslängden för en komponent, λ kallas då felintensitet. Exponentialfördelningen är minneslös, det vill säga,

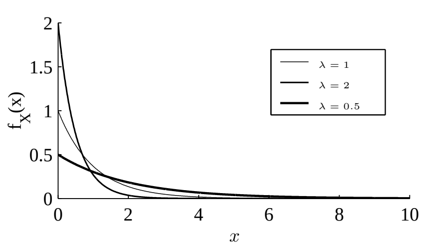
$$\mathsf{P}(X \le t + T | X > t) = \mathsf{P}(X \le T).$$

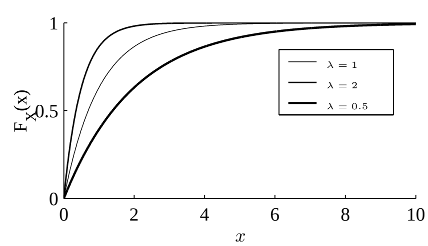
I ord betyder detta att den återstående livslängden är oberoende av det förflutna — komponenter åldras eller slits inte.

$$\mathsf{E}[X] = \frac{1}{\lambda} \qquad\qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

Intressant specialfall:

$$X_1 \sim \exp(\lambda_1),\ X_2 \sim \exp(\lambda_2) \text{ och oberoende } \Longrightarrow \min(X_1, X_2) \sim \exp(\lambda_1 + \lambda_2).$$





Negativ binomialfördelning
är en generalisering av den geometriska fördelningen ovan. X är negativt binomialfördelad med parametrar p och r, om

$$\mathsf{P}\left(X=k\right)=\binom{k-1}{r-1}(1-p)^{k-r}p^r,\quad k=r,r+1,r+2,\ldots$$

Beskriver antalet försök vi måste göra tills vi lyckas r:te gången, där varje försök lyckas med sannolikheten p. Med det inser vi att detta är fördelningen för en summa av r stycken oberoende geometriskt fördelade stokastiska variabler, och då $r=1$ får vi precis geometriska fördelningen.

$$\mathsf{E}\left[X\right]=\frac{r}{p}\qquad\qquad \mathrm{Var}\left(X\right)=r\frac{1-p}{p^2}.$$

Γ-fördelningen (Gamma-fördelningen)
X är Γ-fördelad med parametrar n och λ om

$$f_X(x)=\frac{\lambda^n}{\Gamma(n)}x^{n-1}\mathrm{e}^{-\lambda x},\quad x\geq 0,$$

Funktionen i nämnaren $\Gamma(n)$ definieras av

$$\Gamma(n)=\int_0^\infty z^{n-1}\mathrm{e}^{-z}\,dz,\quad n\geq 0.$$

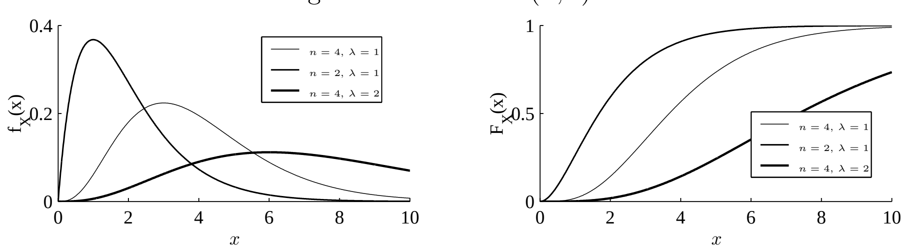
Om n är heltalig så är $\Gamma(n)=(n-1)!$, och fördelningen $f_X(x)$ är fördelningen för en summa av n oberoende, exponentialfördelade stokastiska variabler, alla med parameter λ. För $n=1$ är Γ-fördelningen lika med exponentialfördelningen.

$$\mathsf{E}\left[X\right]=\frac{n}{\lambda}\qquad\qquad \mathrm{Var}\left(X\right)=\frac{n}{\lambda^2}.$$

Intressant relation:

$$X\sim\Gamma(n,\lambda)\Longrightarrow\lambda X\sim\Gamma(n,1),$$

där vi kan få konfidensintervall för λ genom att avläsa $\Gamma(n,1)$-tabeller.



Binomialfördelningen

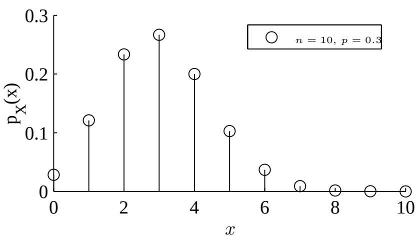
X är binomialfördelad med parametrar n och p, $X \sim \mathrm{Bin}(n, p)$ om

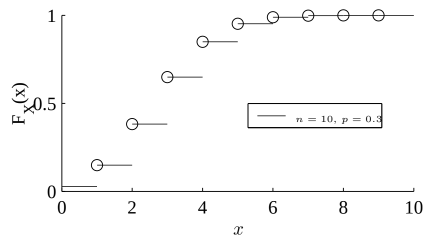
$$\mathsf{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \ldots, n.$$

Typisk situation är att X är antalet lyckade försök i en serie av n oberoende försök, vardera med sannolikhet p att lyckas.

$$\mathsf{E}[X] = np \qquad\qquad \mathrm{Var}(X) = np(1-p).$$

Om n är stor och p liten, kan binomialfördelningen approximeras med Poissonfördelningen med $c = np$.





Poisson-fördelningen

X är Poissonfördelad med parameter c, $X \sim \mathrm{Po}(c)$, om

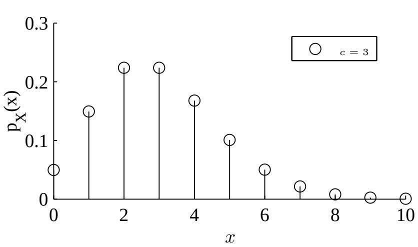
$$\mathsf{P}(X = k) = \frac{c^k}{k!} \mathrm{e}^{-c}, \quad k = 0, 1, \ldots$$

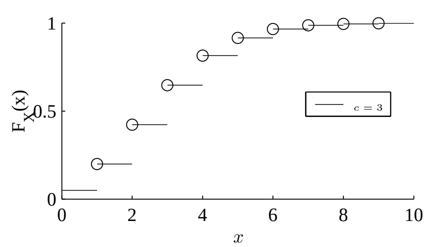
Poissonfördelningen räknar antalet händelser, typiskt över ett tidsintervall T, där händelser sker med intensitet λ. En vanlig modell för antalet inkommande samtal till en telefonstation under tiden T. Parametern c tolkas som det förväntade antalet händelser, och uppfyller $c = \lambda T$. Tiden mellan två händelser är exponentialfördelad med parameter λ, och på grund av minneslösheten, är *antalet* händelser på disjunkta (ej överlappande) intervall *oberoende*.

$$\mathsf{E}[X] = c \qquad\qquad \mathrm{Var}(X) = c$$

Intressant relation:

$$X_1 \sim \mathrm{Po}(c_1), X_2 \sim \mathrm{Po}(c_2) \text{ och oberoende} \implies X_1 + X_2 \sim \mathrm{Po}(c_1 + c_2).$$





Hypergeometrisk fördelning
X är hypergeometriskt fördelad med parametrar s, v och n, om

$$\mathsf{P}\,(X = k) = \frac{\binom{s}{k}\binom{v}{n-k}}{\binom{s+v}{n}}, \quad k = \max(n-v, 0), \ldots, \min(s, n).$$

Detta är ingen fördelning att minnas egentligen, utan den kan lätt härledas när den behövs. Fördelningen uppstår då vi vill plocka n kulor (utan återläggning) ur en urna med s svarta och v vita kulor, och sedan räknar antalet svarta bland dessa n dragna. Vanligtvis brukar problemen ställas i form av defekta och hela enheter i ett parti varor istället för svarta och vita kulor.
För kompletthetens skull infogas här också väntevärdet och variansen för den hypergeometriska fördelningen ovan.

$$\mathsf{E}\,[X] = \frac{s}{s+v} \cdot n \qquad \mathrm{Var}\,(X) = \frac{sv}{(s+v)^2} \cdot n \cdot \frac{s+v-n}{s+v-1}.$$

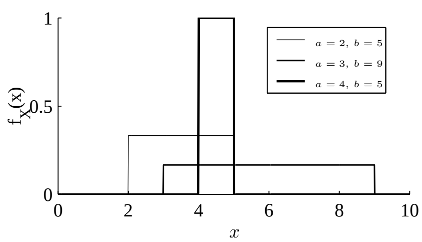
Likformig fördelning (kontinuerliga fallet)
X är likformigt fördelad på intervallet $[a, b]$, $X \sim \mathrm{Likf}(a, b)$, om

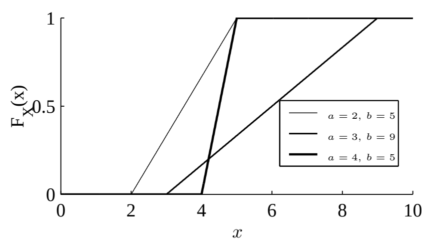
$$f_X(x) = \frac{1}{b-a}, \ a \le x \le b \qquad F_X(x) = \frac{x-a}{b-a}, \ a \le x \le b.$$

Detta är en grundläggande fördelning och definierar vad vi menar med ”på måfå”, och såsom brukligt för grundläggande saker är beteckningen $\mathrm{Likf}(a, b)$ ej standard. Alla skriver olika.

$$\mathsf{E}\,[X] = \frac{a+b}{2} \qquad \mathrm{Var}\,(X) = \frac{(b-a)^2}{12}.$$

Intressant relation: Vi vet att funktioner av stokastiska variabler i sig är nya stokastiska variabler. Om Y godtycklig stokastisk variabel med fördelningsfunktion $F_Y(t)$, så är $X = F_Y(Y)$ en ny stokastisk variabel och $X \sim \mathrm{Likf}(0, 1)$.





Likformig fördelning (diskreta fallet)
X är likformigt fördelad på talen $\{a, a+1, \ldots, b\}$, om

$$\mathsf{P}\,(X = k) = \frac{1}{b-a+1}, \quad a \le k \le b.$$

$$\mathsf{E}\,[X] = \frac{a+b}{2} \qquad \mathrm{Var}\,(X) = \frac{(b-a)(b-a+2)}{12}.$$

Bernoullifördelning (tvåpunktsfördelning)
X är Bernoullifördelad med parameter p om den bara kan anta värdena 0 och 1, och

$$\mathsf{P}(X=1)=p \qquad\qquad \mathsf{P}(X=0)=1-p.$$

Ett annat vanligt namn på en sådan variabel är indikatorvariabel.

$$\mathsf{E}[X]=p \qquad\qquad \mathrm{Var}(X)=p(1-p)$$

En summa av n oberoende Bernoullivariabler är binomialfördelad med parametrar n och p.

Normalfördelningen
X är normalfördelad med parametrar μ och σ^2, $X \sim \mathrm{N}(\mu, \sigma^2)$ om

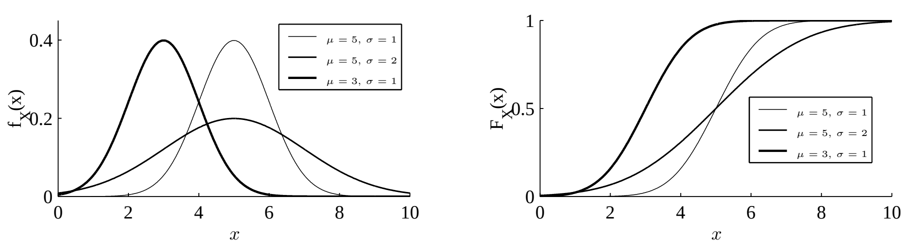
$$f_X(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}.$$

Denna dyker upp precis överallt! Under väldigt allmänna villkor kommer summor av oberoende stokastiska variabler att vara approximativt normalfördelade.

$$\mathsf{E}[X]=\mu \qquad\qquad \mathrm{Var}(X)=\sigma^2.$$

Observera att en del använder beteckningssättet $\mathrm{N}(\mu, \sigma)$ och anger standardavvikelsen istället för variansen. **Alla linjärkombinationer av oberoende normalfördelade stokastiska variabler är i sig normalfördelade.** Grundläggande relation:

$$X \sim \mathrm{N}(\mu, \sigma^2) \implies \frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1).$$



χ^2-fördelningen ("Tji-två", eng. Chi-square)
Förekommer om $X_1, \ldots, X_n$ är oberoende $\mathrm{N}(0,1)$ så är $Y = X_1^2 + \cdots + X_n^2$, χ^2-fördelad med parameter n, vilket utläses "med n frihetsgrader". Speciellt: $Y_1 \sim \chi^2_m$, $Y_2 \sim \chi^2_n$ och oberoende, ger $Y_1 + Y_2 \sim \chi^2_{n+m}$.

$$\mathsf{E}[Y]=n \qquad\qquad \mathrm{Var}(Y)=2n.$$

χ^2-fördelningen är ett specialfall av Γ-fördelningen med $(n, \lambda) = (n/2, 1/2)$.