TDAB01 Sannolikhetslära och Statistik

Jose M. Peña
IDA, Linköpings Universitet

Föreläsning 4

Översikt

- **Täthetsfunktion**
- **Likformig fördelning**
- **Exponentialfördelningen**
- **Gammafördelningen**
- **Betafördelningen**
- **Normalfördelningen**
- **t-fördelningen**

Kontinuerliga slumpvariabler

- Kontinuerliga slumpvariabler kan anta alla reella värden på ett intervall (a, b), speciellt $(-\infty, \infty)$.
- X kontinuerlig $\Rightarrow P(x) = 0$ för alla x. Pmf **inte** användbar.
- Fördelningsfunktionen **funkar** dock: $F(x) = \boldsymbol{P}(X \leq x)$.
- Eftersom $P(x) = 0$ för alla x, så gäller $\boldsymbol{P}(X \leq x) = \boldsymbol{P}(X < x)$.
- Om X kontinuerlig slumpvariabel: $F(x)$ **kontinuerlig**. Inga hopp. **Icke-avtagande**.

$$\lim_{x \to -\infty} F(x) = 0 \text{ och } \lim_{x \to +\infty} F(x) = 1$$

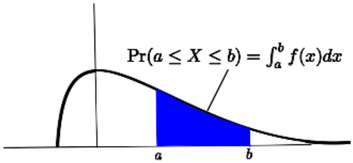
Täthetsfunktion

Definition. **Täthetsfunktionen** $f(x)$ för en kontinuerlig slumpvariabel X är derivatan av cdf:en, dvs

$$f(x) = F'(x)$$

- Täthetsfunktion kallas ofta **pdf** (probability density function).
- cdf:en $F(x)$ är antiderivatan av pdf:en.
- Sannolikheter för intervall ges av **ytor under pdf:en**

$$\boldsymbol{P}(a < X < b) = \int_a^b f(x)dx = F(b) - F(a)$$



- Täthetsfunktioner integrerar till ett

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

Väntevärde och varians

- Täthetsfunktionens värden, t ex $f(2)$, är **inte** en sannolikhet. $f(2) > 1$ helt ok. Men $f(x) \geq 0$ måste gälla.
- För litet ϵ: $\boldsymbol{P}\left(a - \frac{\epsilon}{2} \leq X \leq a + \frac{\epsilon}{2}\right) \approx \epsilon \cdot f(a)$
- Se Example 4.1 i Baron.
- För diskreta slumpvariabler

$$\mathbb{E}(X) = \mu = \sum_x x \cdot P(x) \text{ och } Var(X) = \mathbb{E}\left[(X-\mu)^2\right] = \sum_x (x-\mu)^2 P(x)$$

- För kontinuerliga slumpvariabler

$$\mathbb{E}(X) = \mu = \int x \cdot f(x) dx \text{ och } Var(X) = \mathbb{E}\left[(X-\mu)^2\right] = \int (x-\mu)^2 f(x) dx$$

Simultanfördelning för kontinuerliga variabler

- **Simultan fördelningsfunktion**

$$F_{(X,Y)}(x,y) = \boldsymbol{P}(X \leq x \cap Y \leq y)$$

- **Simultan täthetsfunktion**

$$f_{(X,Y)}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{(X,Y)}(x,y)$$

- Ofta skriver man bara $f(x,y)$ istället för $f_{(X,Y)}(x,y)$.
- **Kovarians**

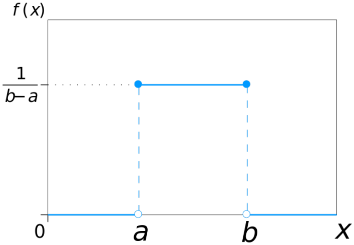
$$\begin{aligned} Cov(X,Y) &= \mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right] \\ &= \int\int (X-\mu_X)(Y-\mu_Y) f(x,y) dxdy \end{aligned}$$

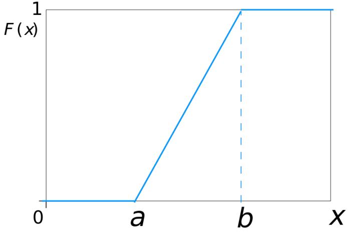
Likformig fördelning

Definition. En **likformig fördelad** variabel, dvs $X \sim U(a,b)$, har täthetsfunktionen

$$f(x) = \frac{1}{b-a} \quad \text{för } a \le x \le b, \text{ och } f(x) = 0 \text{ annars}$$

- Likformig heter **U**niform på engelska.





- Obs. att en variabel kan inte vara likformig fördelad i intervallet $(-\infty, \infty)$.

Likformig fördelning

- **Väntevärde**:

$$\mathbb{E}(X) = \int x \cdot f(x)dx = \frac{1}{b-a}\int xdx = \frac{1}{b-a}\left[\frac{1}{2}x^2\right]_a^b$$

$$= \frac{1}{2(b-a)}\left(b^2 - a^2\right) = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}$$

- **Varians**: $Var(X) = \mathbb{E}(X^2) - \mu^2$

$$\mathbb{E}(X^2) = \int x^2 \cdot f(x)dx = \frac{1}{b-a}\int x^2 dx = \frac{a^2+b^2+ab}{3}$$

$$Var(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{a^2+b^2+ab}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$$

- Alternativ härledning, se Baron sida 81. Alla likformiga variabler kan genereras från **standardmedlemmen**: $Y \sim U(0,1)$ genom följande resultat

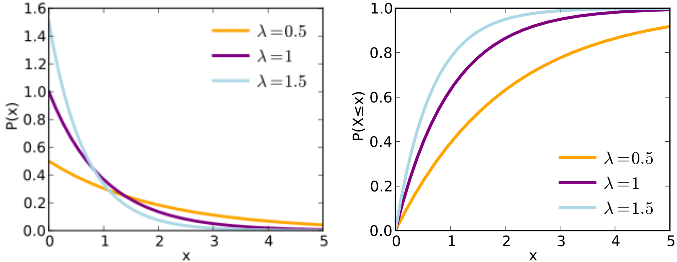
$$X = a + (b-a)Y \text{ där } Y \sim U(0,1) \Longrightarrow X \sim U(a,b)$$

Exponentialfördelningen

Definition. En **exponentialfördelad** variabel, dvs $X \sim Exp(\lambda)$, har täthetsfunktionen

$$f(x) = \lambda e^{-\lambda x} \text{ för } x > 0$$

- Egenskaper av $X \sim Exp(\lambda)$
  - $\mathbb{E}(X) = \frac{1}{\lambda}$
  - $Var(X) = \frac{1}{\lambda^2}$
  - $F(x) = \int_0^x f(y)dy = 1 - e^{-\lambda x}$ eftersom $(e^{g(z)})' = e^{g(z)}g'(z)$ och $\lim_{x\to\infty} F(x) = 1$.
- Parametern λ betyder det samma som i en Poisson fördelning, dvs förväntat antal händelser i en tidsperiod. Men medan en Poisson fördelning modellerar **antal** händelser, en exponentialfördelning modellerar **tiden** mellan två händelser, s k **interarrival** tiden. Se Example 4.5 i Baron.



Exponentialfördelningen

- Exponentialfördelningen används mest för att modellera **tiden**, eftersom tiden mellan två Poissonhändelser är exponentialfördelad.
- **Bevis**. Låt $Po(\lambda t)$ räkna antalet händelser i tidsintervallet $[0, t]$. Låt T vara tiden till nästa händelse

$$\begin{aligned}\boldsymbol{P}(T \leq t) &= 1 - \boldsymbol{P}(T > t) \\ &= 1 - \boldsymbol{P}(\text{inga händelser i intervallet } [0, t]) \\ &= 1 - \frac{e^{-\lambda t}(\lambda t)^0}{0!} = 1 - e^{-\lambda t}\end{aligned}$$

  vilket är cdf:en för en $Exp(\lambda)$ variabel.
- Exponentialfördelade variabler är **minneslösa**:

$$\boldsymbol{P}(T > t | T > s) = \frac{\boldsymbol{P}(T > t)}{\boldsymbol{P}(T > s)} = \frac{1 - \boldsymbol{P}(T \leq t)}{1 - \boldsymbol{P}(T \leq s)} = \frac{e^{-\lambda t}}{e^{-\lambda s}} = e^{-\lambda(t-s)} = \boldsymbol{P}(T > t-s)$$

  for $t > s \geq 0$, dvs bara längden mellan s och t spelar roll.
- Geometriska fördelningen är också minneslös. Den är faktiskt den diskreta motsvarande av exponentialfördelningen.

Gammafördelningen

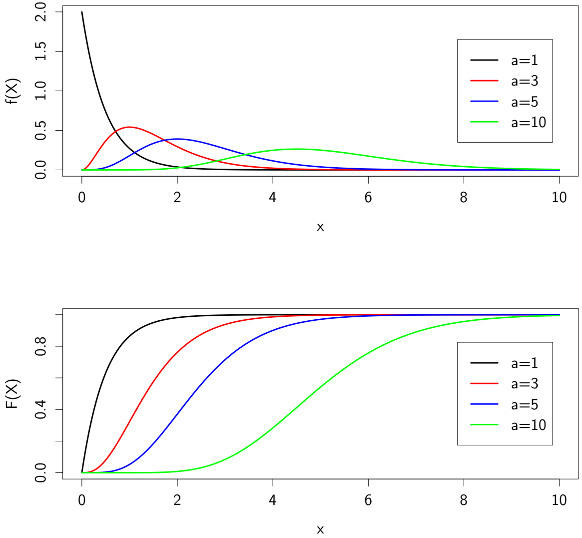
- Antag att tiden för att ladda ner en fil är $Exp(\lambda)$ fördelad. Tiden för att ladda ner α filer följer en $Gamma(\alpha, \lambda)$ fördelning om nedladdningstiderna är oberoende.

Definition. Om $X_1, X_2, \ldots, X_\alpha$ är α stycken **oberoende** $Exp(\lambda)$ variabler, då

$$Y = X_1 + X_2 + \ldots + X_\alpha \sim Gamma(\alpha, \lambda)$$

- α kallas för en **shape**parameter, och λ är en **frekvens**parameter.
- $Exp(\lambda) = Gamma(1, \lambda)$
- Egenskaper av $X \sim Gamma(\alpha, \lambda)$
  - $\mathbb{E}(X) = \frac{\alpha}{\lambda}$
  - $Var(X) = \frac{\alpha}{\lambda^2}$
- α och λ kan faktiskt anta vilka positiva reella värden som helst. Då, funkar inte den tolkingen ovanpå, men ovanpå väntevärdet och variansen stämmer ändå.

Gammafördelningen



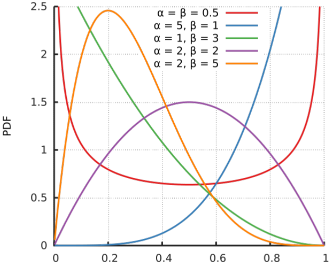
Betafördelningen

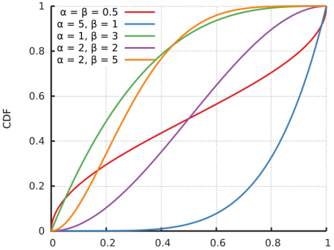
Definition. En **betafördelad** variabel, dvs $X \sim Beta(\alpha, \beta)$, har täthetsfunktionen

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad \text{för } 0 \leq x \leq 1$$

där $B(,)$ är **betafunktionen**.

- Egenskaper av $X \sim Beta(\alpha, \beta)$
  - $\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}$
  - $Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
- Passar kontinuerliga variabler i intervallet $[0, 1]$, t ex **andelar** eller **sannolikheter**.





Normalfördelningen

- Mest använda fördelningen, inte minst pga den centrala gränsvärdessatsen (Fö5).

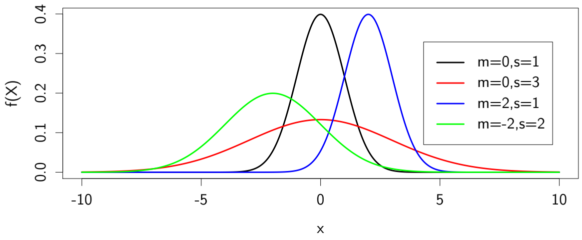
Definition. En **normalfördelad** variabel, dvs $X \sim N(\mu, \sigma^2)$, har täthetsfunktionen

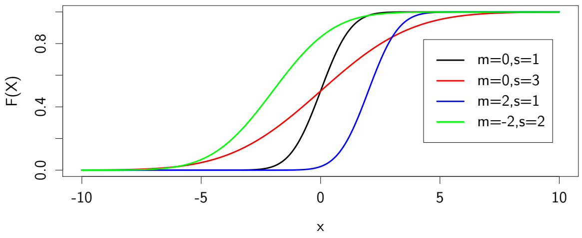
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad \text{för} \quad -\infty < x < \infty$$

- Egenskaper av $X \sim N(\mu, \sigma^2)$
  - $\mathbb{E}(X) = \mu$
  - $Var(X) = \sigma^2$
- cdf:en finns inte i sluten form. Om $Z \sim N(0,1)$ så är cdf:en

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz$$

Normalfördelningen





Normalfördelningen

- **Standardmedlem**: $Z \sim N(0,1)$

$$X = \mu + \sigma Z \text{ där } Z \sim N(0,1) \Longrightarrow X \sim N(\mu, \sigma^2)$$

- **Standardisering**

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

- Standardisering är praktiskt. Låt $X \sim N(\mu = 900, \sigma = 200)$

$$\begin{aligned} \boldsymbol{P}(600 < X < 1200) &= \boldsymbol{P}\Big(\frac{600-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{1200-\mu}{\sigma}\Big) \\ &= \boldsymbol{P}(-1.5 < Z < 1.5) \\ &= \Phi(1.5) - \Phi(-1.5) = 0.9332 - 0.0668 = 0.8664 \end{aligned}$$

- Se Examples 4.11 och 4.12 i Baron.

t-fördelningen

- Normalfördelningen har **tunna svansar**. Mycket osannolikt att observera extrema observationer.
- t-fördelningen är en generalisering av normalfördelningen med en parameter ν (**frihetsgrader**) som modellerar hur tunga svansarna är.

Definition. En ***t*-fördelad** variabel, dvs $X \sim t_\nu(\mu, \sigma^2)$, har täthetsfunktionen

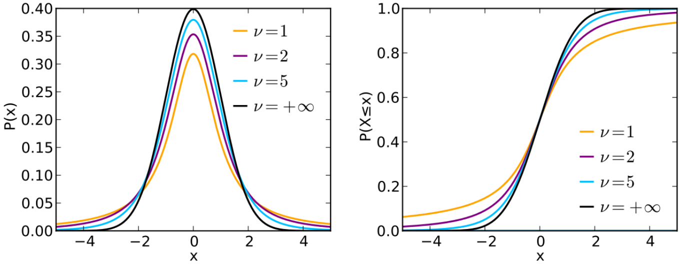
$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu\sigma^2}}\left(1 + \frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}} \quad \text{för } -\infty < x < \infty$$

där $\Gamma()$ är **gammafunktionen**.

- Egenskaper av $X \sim t_\nu(\mu, \sigma^2)$
  - $\mathbb{E}(X) = \mu$ om $\nu > 1$, odefinerad annars
  - $Var(X) = \sigma^2 \frac{\nu}{\nu-2}$ om $\nu > 2$, $Var(X) = \infty$ om $1 < \nu \leq 2$, odefinerad annars
- $t_\nu(0, 1)$ är standardmedlemmen. Samma standardisering som förut.

t-fördelningen

- Normalfördelningen fås när $\nu \to \infty$.



- Viktig koppling mellan $t-$fördelning och normalfördelning:
  - $X_1, .., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. σ^2 **känd**. $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
  - $X_1, .., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. σ^2 **okänd**, skattas med s^2. $T = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}(0,1)$.
- Vi återkommer till detta koppling senare i kursen.

Översikt

- **Täthetsfunktion**
- **Likformig fördelning**
- **Exponentialfördelningen**
- **Gammafördelningen**
- **Betafördelningen**
- **Normalfördelningen**
- **t-fördelningen**