TDAB01 Probability and Statistics

Maryna Prus
IDA, Linköping University

Lecture 4: Families of Continuous Distributions

Overview

- **Continuous random variables**
- **Uniform distribution**
- **Exponential distribution**
- **Normal distribution**
- **Gamma distribution**
- **Beta distribution**
- **t-distribution**

Continuous random variables

- Continuous random variables can take all values on an interval (a, b), especially $(0, \infty)$, $(-\infty, \infty)$
- X continuous $\Rightarrow P(x) = 0$ for all $x \Rightarrow$ pmf **not** useful
- The distribution function (cdf) however **works**: $F(x) = \boldsymbol{P}(X \leq x)$
- Since $P(x) = 0$ for all x, then $\boldsymbol{P}(X \leq x) = \boldsymbol{P}(X < x)$
- If X continuous random variable, $F(x)$ **continuous** and

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

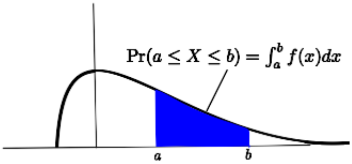
Density function

Definition. **(Probability) density function** $f(x)$ of a continuous random variable X is the derivative of the cdf of X

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

- Density function is often called **pdf** (probability density function)
- Cdf $F(x)$ is antiderivative of pdf $f(x)$
- Interval probabilities are given by **areas under the pdf**

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



- $\boldsymbol{P}(a < X < b) = \boldsymbol{P}(a \le X < b) = \boldsymbol{P}(a < X \le b) = \boldsymbol{P}(a \le X \le b)$

Expected value and variance

- Integral of pdf on $(-\infty, \infty)$ is 1:

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

- Values of pdf, for example $f(2)$, are **not** a probabilities: $f(2) > 1$ ok
  But $f(x) \geq 0$ must hold
- Example: Example 4.1 in textbook
- For discrete random variables:

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

- For continuous random variables:

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

Joint distribution for continuous random variables

- **Joint distribution function**

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

- **Joint density function**

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

- **Kovariance** between X and Y

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

- X and Y independent. Then

$$Cov(X,Y) = 0 \quad \& \quad f_{(X,Y)}(x,y) = f_X(x)f_Y(y), \quad \text{for all } x,y$$

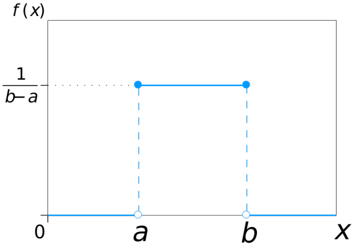
Uniform distribution

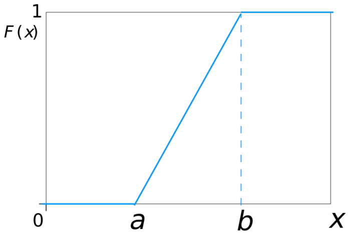
Definition. A random variable X is **uniformly distributed** on (a, b) if pdf of X is given by

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b, \\ 0, & \text{otherwise.} \end{cases}$$

- Notation: $X \sim U(a, b)$
- Cdf of $X \sim U(a, b)$

$$F(x) = \begin{cases} 0, & x \leq a, \\ \frac{x-a}{b-a}, & a < x < b, \\ 1, & x \geq b. \end{cases}$$





Uniform distribution

- Uniform distribution on intervals $(-\infty, \infty)$, (a, ∞) or $(-\infty, b)$ impossible
- **Expected value**:

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

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

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

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx = \frac{1}{b-a}\int_{-\infty}^{\infty} x^2 dx = \frac{\left[x^3\right]_a^b}{3(b-a)} = \frac{a^2+b^2+ab}{3}$$

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

- $Y \sim U(0,1)$ - **Standard Uniform distribution**
- For $X \sim U(a,b)$ and $Y \sim U(0,1)$ holds

$$X = a + (b-a)Y$$

Exponential distribution

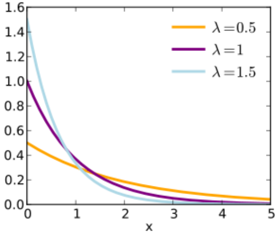
Definition. A random variable X is **exponentially distributed** with parameter $\lambda > 0$, i.e. $X \sim Exp(\lambda)$, if pdf of X is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

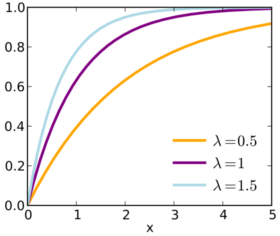
- Cdf of $X \sim Exp(\lambda)$

$$F(x) = \begin{cases} 0, & x \leq 0, \\ 1 - e^{-\lambda x}, & x > 0. \end{cases}$$

$f(x)$



$F(x)$



Exponential distribution

- Exponential distribution usually models **time**
- For example
  - time between two rare events (Poisson distribution models number of rare events)
  - time until some specific event occurs
  - duration
- Expected value and variance:

$$\mathbb{E}(X) = \frac{1}{\lambda} \quad \& \quad Var(X) = \frac{1}{\lambda^2}$$

- λ - frequency parameter, number of events per time unit

Normal distribution

Definition. A random variable X has **normal distribution** with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, i. e. $N(\mu, \sigma^2)$, if pdf of X is given by

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

- Expected value and variance of $X \sim N(\mu, \sigma^2)$

$$\mathbb{E}(X) = \mu \quad \& \quad Var(X) = \sigma^2$$

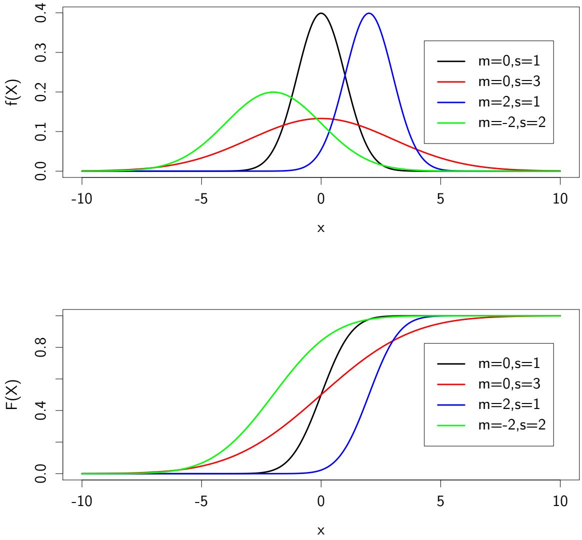
- CDF has no closed form
- For $Z \sim N(0,1)$ cdf:

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

- Values of $\Phi(z)$ are given in tables
- $Z \sim N(0,1)$ - **standard normal distribution**
- Normal distribution is also called Gaussian distribution

Normal distribution

$m = \mu,\ s = \sigma^2$



Normal distribution

- Z standard nomally distributed: $Z \sim N(0,1)$

$$\Rightarrow \quad X = \mu + \sigma Z \sim N(\mu, \sigma^2)$$

- **Standardization** of $X \sim N(\mu, \sigma^2)$

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

- Example: $X \sim N(\mu = 900, \sigma = 200)$

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

- See Examples 4.11 and 4.12 in textbook

Gamma distribution

Definition. A random variable X has **Gamma distribution** with parameters $\alpha > 0$ and $\lambda > 0$ if pdf of X is given by

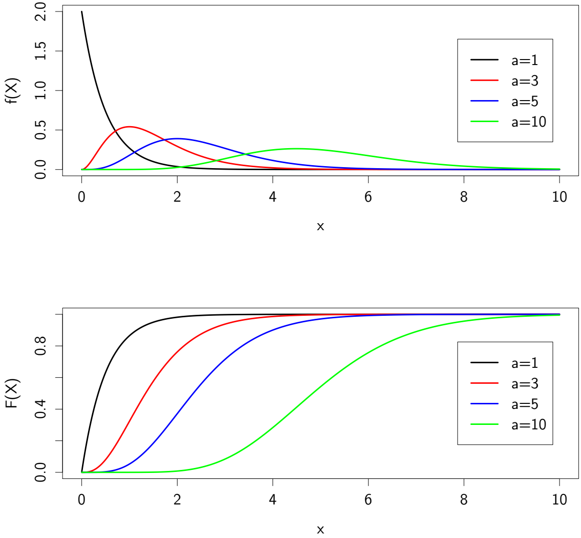
$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0, \\ 0, & \text{otherwise,} \end{cases}$$

where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ is **Gamma function**.

- Notation: $X \sim Gamma(\alpha, \lambda)$ or $X \sim Ga(\alpha, \lambda)$
- For discrete α: $\Gamma(\alpha) = \alpha!$
- α - shape parameter, λ - frequency parameter
- $Exp(\lambda) = Gamma(1, \lambda)$
- Expected value and variance of $X \sim Gamma(\alpha, \lambda)$:
  - $\mathbb{E}(X) = \frac{\alpha}{\lambda}$
  - $Var(X) = \frac{\alpha}{\lambda^2}$
- For discrete α and independent $Exp(\lambda)$ distributed $X_1, X_2, \ldots, X_\alpha$ holds

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

Gamma distribution



$a = \alpha, \ \lambda = 2$

Poisson-Gamma formula

- $T \sim Gamma(\alpha, \lambda)$, $\alpha, \lambda > 0$, α integer

  $\Rightarrow$ T - time before α-th rare event

  $\Rightarrow$ $P(T > t)$ - probability that α-th rare event occurs after time t

- $X \sim Po(t\lambda)$, $t, \lambda > 0$

  $\Rightarrow$ X - number of rare events that occur before time t

  $\Rightarrow$ $P(X < \alpha)$ - probability that less than α rare events occur before time t

Poisson-Gamma formula:

$$P(T > t) = P(X < \alpha)$$

and

$$P(T \leq t) = P(X \geq \alpha)$$

Beta distribution

Definition. A random variable X has **Beta distribution** with parameters $\alpha > 0$ and $\beta > 0$, i. e. $X \sim Beta(\alpha, \beta)$, if pdf of X is given by

$$f(x) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & 0 < x < 1, \\ 0, & \text{otherwise}, \end{cases}$$

where $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is **Beta function**.

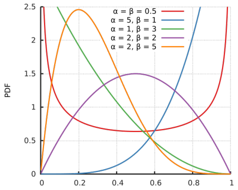
- Expected value and variance of $X \sim Beta(\alpha, \beta)$:

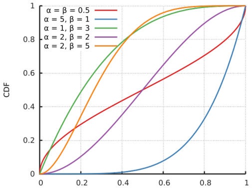
$$\mathbb{E}(X) = \frac{\alpha}{\alpha + \beta} \quad \& \quad Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

- $X \sim Gamma(\alpha, \lambda)$ and $Y \sim Gamma(\beta, \lambda)$ independent

$$\Rightarrow \quad Z = X/(X+Y) \sim Beta(\alpha, \beta)$$

- Beta distribution fits continuous variables in range $[0, 1]$, e. g. proportions





t-distribution

Definition. A random variable X has ***t*-distribution** with ν degrees of freedom if pdf of X is given by

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, -\infty < x < \infty$$

where $\Gamma()$ is **Gamma function**.

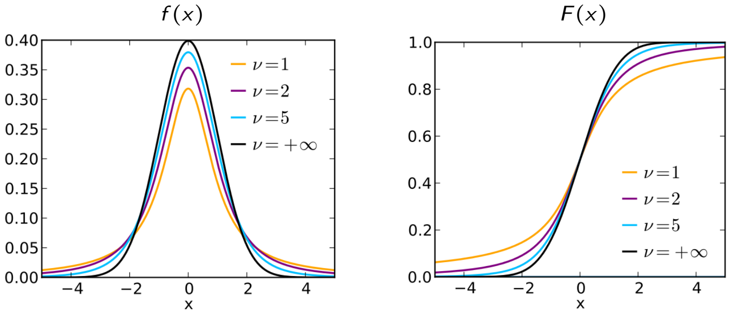
- Notation: $X \sim t_\nu$ or $X \sim t(\nu)$ or $X \sim T(\nu)$
- Expected value and variance of $X \sim t(\nu)$

$$\mathbb{E}(X) = 0, \nu > 1 \quad \& \quad Var(X) = \frac{\nu}{\nu-2}, \nu > 2$$

- Connection between t-distribution and normal distribution:
  - $X_1, .., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, σ^2 **known** $\Rightarrow Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
  - $X_1, .., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, σ^2 **unknown**, estimator $s^2 \Rightarrow T = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$

t-distribution

- $t(\nu) \to N(0,1)$ if $\nu \to \infty$:



- Normal distribution has **thin tails**
- t-distribution models heavy tails

Thank you for your attention!