Probability and Statistics

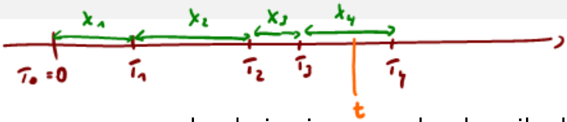
4 – Continuous Random Variables

Stefan Heiss

Technische Hochschule Ostwestfalen-Lippe
Dep. of Electrical Engineering and Computer Science

December 22, 2023

Poisson Processes



Definition (4.67)

A process in which discrete similar events occur randomly in time, can be described by the following sequences of random variables:

- *Arrival times*: $T = (T_1, T_2, \dots)$ (also set $T_0 := 0$)
- *Inter-arrival times*: $X = (X_1, X_2, \dots)$ with $X_n := T_n - T_{n-1}$

$$\left(\implies \quad T_n = \sum_{i=1}^{n} X_i \text{ for all } n \in \mathbb{N} \right)$$

Such a process can also be described with the following set of random variables:

- *Counting process*: $N = \{N_t \mid t \geq 0\}$ with $N_t := \max\{n \in \mathbb{N}_0 \mid T_n \leq t\}$

$$\left(\implies \quad T_n = \min\{t \geq 0 \mid N_t = n\} \text{ for all } n \in \mathbb{N}_0 \right)$$

Poisson Processes

Lemma (4.68)

The following are equivalent:

(i) *At least n arrivals occurred in the interval* $(0, t]$.

(ii) $N_t \geq n$

(iii) $T_n \leq t$

Renewal and Poisson Processes

Definition (4.69)

- A process in which events occur randomly in time is called a *renewal process*, if the inter-arrival times $X_1, X_2, \ldots$ are independent, identically distributed random variables.
- A renewal process is called a *Poisson process*, if it satisfies the *strong* renewal assumption, that at each fixed time, the process restarts probabilistically, independent of the past.

$\Leftrightarrow \quad X_i \sim exp(\lambda)$

Poisson Processes

Theorem (4.70)

Given a Poisson process with arrival times $T = (T_0 = 0, T_1, T_2, \dots)$, *inter-arrival times* $X = (X_1, X_2, \dots)$ *and counting variables* $N = \{N_t \mid t \geq 0\}$, *there exists some parameter* $\lambda > 0$, *such that:*

- $X_n \sim \exp(\lambda)$ *for all* $n \in \mathbb{N}$
- $T_n \sim \mathit{Erlang}(n, \lambda)$ *for all* $n \in \mathbb{N}$
- $N_t \sim \mathrm{Poisson}(\lambda t)$ *for all* $t > 0$

Def. of Poisson process

$T_n = X_1 + \dots + X_n \sim \text{Erlang}(n, \lambda)$

Poisson Processes: $N_t \sim \mathrm{Poisson}(\lambda t)$ for all $t > 0$

$$\Pr(N_t \geq n) \overset{(4.68)}{=} \Pr(T_n \leq t) = F_{T_n}(t) \overset{(4.66)}{=} 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

$$\begin{aligned} \Longrightarrow \qquad \Pr(N_t = n) &= \Pr(N_t \geq n) - \Pr(N_t \geq n+1) \\ &= 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} - \left(1 - \sum_{k=0}^{n} \frac{(\lambda t)^k}{k!} e^{-\lambda t}\right) \\ &= \frac{(\lambda t)^n}{n!} e^{-\lambda t} \end{aligned}$$

$$\Longrightarrow \qquad N_t \sim \mathrm{Poisson}(\lambda t)$$

Poisson Processes

Theorem (4.70)

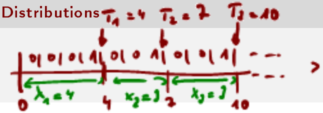
Given a Poisson process with arrival times $T = (T_0 = 0, T_1, T_2, \dots)$, *inter-arrival times* $X = (X_1, X_2, \dots)$ *and counting variables* $N = \{N_t \mid t \geq 0\}$, *there exists some parameter* $\lambda > 0$, *such that:*

- $X_n \sim \exp(\lambda)$ *for all* $n \in \mathbb{N}$
- $T_n \sim \mathit{Erlang}(n, \lambda)$ *for all* $n \in \mathbb{N}$
- $N_t \sim \mathrm{Poisson}(\lambda t)$ *for all* $t > 0$

Definition (4.71)

A Poisson process is said to have *rate* λ, if the inter-arrival times have an exponential distribution with rate parameter λ.

Discrete Time Poisson Processes



Remark (4.72)

A Bernoulli trials process $(B_t)_{t\in\mathbb{N}}$ with independent random variables $B_t \sim \text{Bernoulli}(p)$ can be considered to be a discrete time version of the Poisson process with:

- *Counting process*: $N = (N_t)_{t\in\mathbb{N}_0}$ with

$$N_t \;=\; \sum_{i=1}^{t} B_i \;\sim\; \text{binomial}(t, p)$$

  $N = (N_t)_{t\in\mathbb{N}_0}$ has independent, stationary increments ($N_{t_2} - N_{t_1} = N_{t_2 - t_1}$ for all $t_1, t_2 \in \mathbb{N}_0,\ t_1 \le t_2$)

Discrete Time Poisson Processes

Remark (4.72)

A Bernoulli trials process $(B_t)_{t\in\mathbb{N}}$ with independent random variables $B_t \sim \mathrm{Bernoulli}(p)$ can be considered to be a discrete time version of the Poisson process with:

- *Inter-arrival times*: $X = (X_1, X_2, \dots)$ with independent random variables

$$X_i \sim \mathrm{geometric}(p)$$

- *Arrival times*: $T = (T_1, T_2, \dots)$ has independent, stationary increments and

$$T_i \sim \mathrm{nbino}(i, p) \qquad T_i = X_1 + X_2 + \dots + X_i$$

Chi-Square Distributions

Definition (4.73)

A random variable has a *chi-square distribution with n degrees of freedom* χ^2_n if its pdf is defined by

$$f_X(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \cdot \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

i.e.:

$$\chi^2_n = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$$

$\chi^2_{2n} = \Gamma(n, \frac{1}{2})$

$= \text{Erlang}(n, \frac{1}{2})$

Chi-Square Distributions

Theorem (4.75)

If X has a chi-square distribution with n degrees of freedom, i.e. $X \sim \chi^2_n$, then:

(i) $E(X) = n$

(ii) $Var(X) = 2n$

(iii) $\phi_X(t) = \left(\dfrac{1}{1-2t}\right)^{\frac{n}{2}}$ for $t < 1/2$

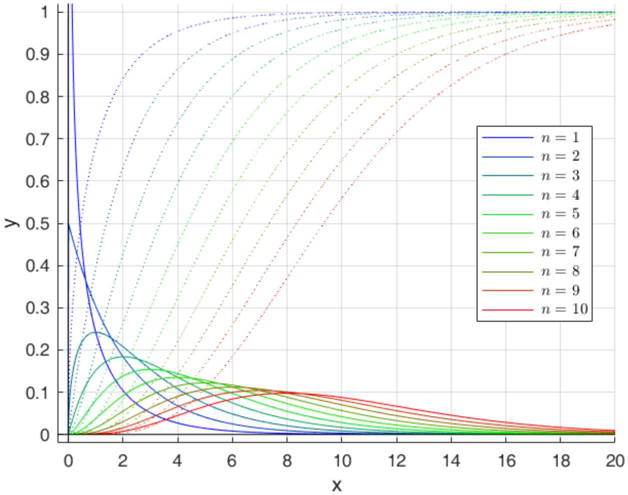
Pf.: (4.52)

Chi-Square Distributions

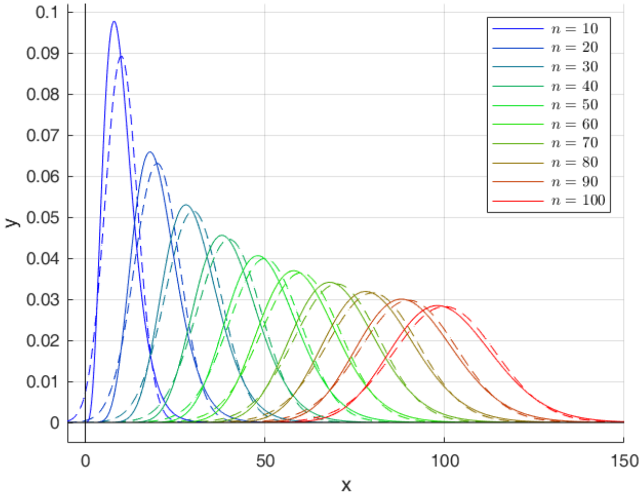
Remark (4.76)

MATLAB provides implementations of pdf's, cdf's and the inverses of cdf's of chi-square distributions under the names `chi2pdf()`, `chi2cdf()` and `chi2inv()`, respectively.

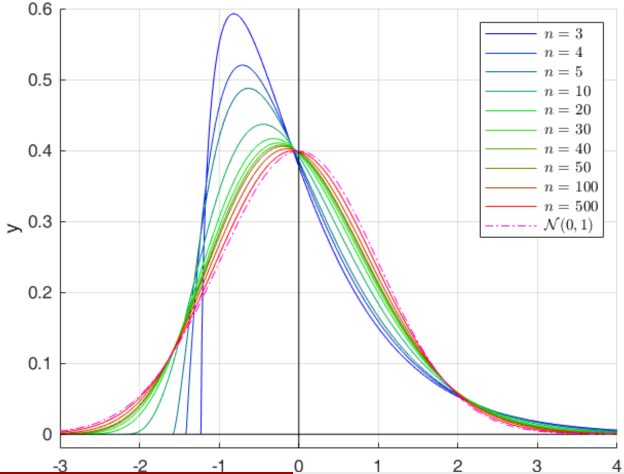
Pdf's and cdf's of $\chi_n^2 = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ distributions for $n = 1, 2, \ldots, 10$



Pdf's of χ^2_n and $\mathcal{N}(n, \sqrt{2n})$ distributions for $n = 10, 20, \ldots, 100$



Pdf's of normalized χ_n^2 distributions for selected values of n and the pdf of the standard normal distribution



Sum of independent random variables $X_i \sim \chi^2_{n_i}$

Theorem (4.77)

If $X_1, \dots, X_m$ are independent random variables with $X_i \sim \chi^2_{n_i}$, then

$$X = X_1 + \cdots + X_m$$

has a chi-square distribution with $X \sim \chi^2_n$, where $n = \sum_{i=1}^m n_i$.

Pf. (4.60) $X_i \sim \Gamma(\frac{n_i}{2}, \frac{1}{2}) \Rightarrow \sum X_i \sim \Gamma(\sum \frac{n_i}{2}, \frac{1}{2})$

$$= \Gamma(\tfrac{1}{2} \sum n_i, \tfrac{1}{2})$$

Sum of squares of independent random variables $Z_i \sim \mathcal{N}(0,1)$

Theorem (4.78)

If $Z_1, \ldots, Z_n$ are independent random variables with $Z_i \sim \mathcal{N}(0,1)$ for all i, then

$$X \;=\; Z_1^2 + \cdots + Z_n^2$$

has a chi-square distribution with n degrees of freedom, i.e. $X \sim \chi_n^2$.

Pf.: $Z \sim \mathcal{N}(0,1) \stackrel{!}{\Rightarrow} Z^2 \sim \chi_1^2$

$\underline{x \geq 0}$: $F_{Z^2}(x) = Pr(Z^2 \leq x) = Pr(-\sqrt{x} \leq Z \leq \sqrt{x}) = \Phi(\sqrt{x}) - \Phi(-\sqrt{x})$

$f_{Z^2}(x) = F'_{Z^2}(x) = \frac{1}{2}x^{-1/2}\cdot\underline{\varphi(\sqrt{x})} + \frac{1}{2}x^{-1/2}\,\underline{\varphi(-\sqrt{x})} = \frac{1}{2}x^{-1/2}\cdot 2\,\underline{\frac{e^{-x/2}}{\sqrt{2\pi}}}$

$= \frac{1}{\sqrt{2\pi}}\, x^{-1/2} e^{-x/2} \leftarrow$ pdf of χ_1^2 $\square$

Sum of squares of independent random variables $Z_i \sim \mathcal{N}(0,1)$

Theorem (4.79)

If $X_1, \ldots, X_n$ are independent random variables with $X_i \sim \mathcal{N}(\mu, \sigma)$ for all i, then

$$\widehat{\sigma_0}^2 = \frac{1}{n}\left((X_1-\mu)^2 + \cdots + (X_n-\mu)^2\right)$$

has a gamma distribution with $\widehat{\sigma_0}^2 \sim \Gamma(\frac{n}{2}, \frac{n}{2\sigma^2})$.

Pf

$$X = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \overset{(4.78)}{\sim} \chi_n^2 = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$$

$$\widehat{\sigma_0}^2 = \frac{\sigma^2}{n} \cdot X \overset{(4.58)(iv)}{\sim} \Gamma\left(\frac{n}{2}, \frac{n}{2\sigma^2}\right)$$

t-Distributions

Definition (4.80)

If Z and Y_n are independent random variables with $Z \sim \mathcal{N}(0,1)$ and $Y_n \sim \chi^2_n$, then the distribution of

$$T_n \quad := \quad \frac{Z}{\sqrt{Y_n/n}}$$

$Pr(Y_n \leq 0) = 0$

is called a *t-distribution with n degrees of freedom*, denoted by:

$$T_n \sim t_n$$

$(n \in \mathbb{N})$

The cdf of a t-distribution with n degrees of freedom is denoted by:

$$F_{t_n}$$

t-Distributions

Theorem (4.81)

If $T_n \sim t_n$, then a pdf of T_n is given by:

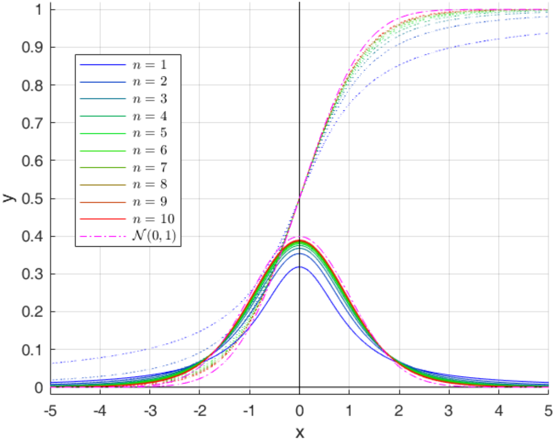
$$f_{t_n}(x) \;=\; \frac{\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}}{\sqrt{n}\cdot \underbrace{B\left(\frac{1}{2},\frac{n}{2}\right)}_{\frac{\Gamma(\frac{1}{2})\Gamma(\frac{n}{2})}{\Gamma(\frac{n+1}{2})}}} \qquad x \in \mathbb{R}$$

Remark (4.82)

The pdf f_{t_n} of a t-distribution is an even function and therefore:

$$F_{t_n}(x) \;=\; 1 - F_{t_n}(-x) \qquad \text{for all } x \in \mathbb{R}$$

Pdf's and cdf's of t_n distributions for $n = 1, 2, \ldots, 10$ and of the standard normal distribution



t-Distributions

Theorem (4.83)

If T_n has a t-distribution $T_n \sim t_n$, then:

(i) $E({T_n}^k)$ *exists, if and only* $k < n$.

(ii) *If* $k < n$ *is odd, then* $E({T_n}^k) = 0$.

(iii) *If* $k < n$ *is even, then:*

$$E({T_n}^k) \;=\; n^{\frac{k}{2}} \cdot \frac{\Gamma\left(\frac{k+1}{2}\right) \cdot \Gamma\left(\frac{n-k}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \cdot \Gamma\left(\frac{n}{2}\right)}$$

k=2

(iv) *If* $n > 2$*, then:* $E(T_n) = 0$ *and* $Var(T_n) = \frac{n}{n-2}$

t-Distributions

Remark (4.84)

MATLAB provides implementations of pdf's, cdf's and the inverses of cdf's of t-distributions under the names `tpdf()`, `tcdf()` and `tinv()`, respectively.

Beta Function

Definition (4.85)

For $x, y > 0$ the *beta function* is defined by:

$$B(x, y) \;:=\; \int_0^1 t^{x-1}(1-t)^{y-1}\,dt$$

Theorem (4.86)

For all $x, y > 0$:

$$B(x, y) \;=\; \frac{\Gamma(x)\cdot\Gamma(y)}{\Gamma(x+y)}$$

Beta Distributions

Definition (4.87)

A random variable has a *beta distribution* $\mathrm{beta}(\alpha_1, \alpha_2)$ for some parameters $\alpha_1, \alpha_2 \in \mathbb{R}^+$, if its pdf is defined by:

$$f_X(x) = \begin{cases} \frac{1}{B(\alpha_1,\alpha_2)} x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1} & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Remark (4.88)

MATLAB provides implementations of pdf's, cdf's and the inverses of cdf's of beta distributions under the names `betapdf()`, `betacdf()` and `betainv()`, respectively.

Beta Distributions

Lemma (4.89)

(i) $$B(\alpha_1, \alpha_2) = \int_0^\infty (1 - e^{-\theta})^{\alpha_1 - 1} e^{-\alpha_2 \theta}\, d\theta$$

(ii) $$B(\alpha_1, \alpha_2) = \int_0^\infty \frac{z^{\alpha_1 - 1}}{(1 + z)^{\alpha_1 + \alpha_2}}\, dz$$

(iii) $$B\left(\frac{1}{2}, \frac{n+1}{2}\right) = 2 \cdot \int_0^{\pi/2} \sin^n(\theta)\, d\theta$$

Beta Distributions

Example (4.90)

If X has a beta distribution $X \sim \mathrm{beta}(\alpha_1, \alpha_2)$, then:

(i) $\alpha_1 = \alpha_2 = 1 \implies X \sim \mathrm{uniform}([0,1])$

(ii) $\alpha_1 = \alpha_2 = 2 \implies f_X(x) = -6(x^2 - x) \cdot I_{(0,1)}$

(iii) $\alpha_1 = \frac{1}{2},\ \alpha_2 = 1 \implies f_X(x) = \frac{1}{2\sqrt{x}} \cdot I_{(0,1)}$

Beta Distributions

Exercise (4.91)

Use (4.86) to provide a new proof of (4.48)(i), i.e.:

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx \;=\; \sqrt{\pi}$$

Beta Distributions

Exercise (4.92)

Prove: If X and Y are independent continuous random variables with $X \sim \Gamma(\alpha_1, \beta)$ and $Y \sim \Gamma(\alpha_2, \beta)$, then a pdf of $Z = \frac{X}{Y}$ is given by:

$$f_Z(z) \quad = \quad \frac{1}{B(\alpha_1, \alpha_2)} \cdot \frac{z^{\alpha_1 - 1}}{(1+z)^{\alpha_1+\alpha_2}} \cdot I_{(0,\infty)}$$