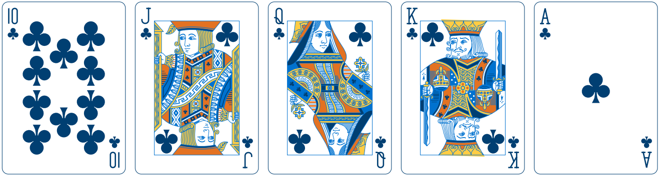
2P7: Probability & Statistics

Manipulating and Combining Distributions

Thierry Savin

Lent 2024



the *royal flush*, the best possible hand in poker, has a probability 0.000154%

UNIVERSITY OF CAMBRIDGE
Department of Engineering

UNIVERSITY OF CAMBRIDGE
Department of Engineering

UNIVERSITY OF CAMBRIDGE
Department of Engineering

In the last lectures:

- We have seen that discrete random variables are described by their probability mass function
- We have seen that continuous random variables are described by their probability density function
- We have given important examples of probability mass and density functions:

Discrete variables
- Bernoulli
- Geometric
- Binomial
- Poisson

Continuous variables
- Exponential
- Gaussian
- Beta

In this lecture, we will manipulate random variables, introduce important transforms of distributions, and see the Central Limit Theorem.

Consider a random variable X, and let $\mathrm{Y} = g(\mathrm{X})$ for some function $g : \mathbb{X} \to \mathbb{Y}$ mapping the support $\mathbb{X}$ of X to the domain $\mathbb{Y}$ of Y.

Can we calculate P_{Y} (or f_{Y} if continuous) from P_{X} (or f_{X})?

We'll start with the case where X is a discrete random variable. Then Y is also discrete. $\mathbb{X} = \{x_1, x_2, \dots\}$ and $\mathbb{Y} = \{y_1, y_2, \dots\}$,

$$
\begin{array}{lcll}
x_1 & \longrightarrow & y_1 & P_{\mathrm{Y}}(y_1) = P_{\mathrm{X}}(x_1) \\
x_2 & \longrightarrow & y_2 & P_{\mathrm{Y}}(y_2) = P_{\mathrm{X}}(x_2) \\
x_3 & \longrightarrow & y_3 & P_{\mathrm{Y}}(y_3) = P_{\mathrm{X}}(x_3) + P_{\mathrm{X}}(x_4) \\
x_4 & \longrightarrow & y_3 & \\
\vdots & & \vdots &
\end{array}
$$

In general,
$$P_{\mathrm{Y}}(y) = \sum_{\{x|g(x)=y\}} P_{\mathrm{X}}(x)$$

If g is *invertible* (one-to-one map between $\mathbb{X}$ and $\mathbb{Y}$), then
$$P_{\mathrm{Y}}(y) = P_{\mathrm{X}}\left(g^{-1}(y)\right)$$

UNIVERSITY OF CAMBRIDGE
Department of Engineering

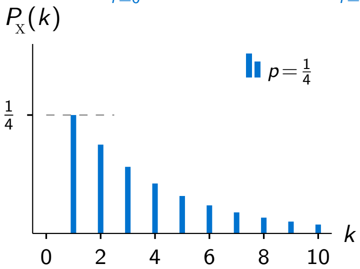
With $\mathrm{X} \sim \mathrm{Geo}(p)$ and $m \in \{1, 2, \dots\}$, what is the PMF of $\mathrm{Y} = (\mathrm{X}-1) \bmod m$? (the modulo operation "$a \bmod b$" returns the remainder of $a \div b$)

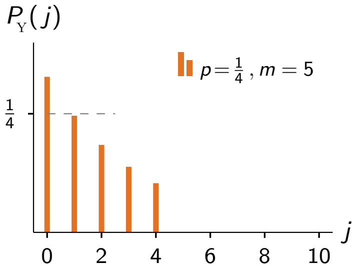
Observe that $\mathbb{Y} = \{0, 1 \dots, m-1\}$ and that

$$
\begin{array}{llllll}
\mathrm{Y} = 0 \text{ if } \mathrm{X} = & 1, & m+1, & 2m+1, & 3m+1\dots \\
\mathrm{Y} = 1 \text{ if } \mathrm{X} = & 2, & m+2, & 2m+2, & 3m+2\dots \\
\quad \dots & & & & \\
\mathrm{Y} = y \text{ if } \mathrm{X} = & \multicolumn{4}{l}{y+1, m+y+1, 2m+y+1, 3m+y+1\dots}
\end{array}
$$

Hence $\{x | g(x) = y\} = \{x = rm + y + 1\}_{r \in \{0,1,\dots\}}$ and we conclude:

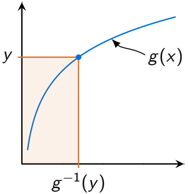
$$P_{\mathrm{Y}}(y) = \sum_{r=0}^{\infty} P_{\mathrm{X}}(rm + y + 1) = \sum_{r=0}^{\infty} p(1-p)^{rm+y} = \frac{p(1-p)^y}{1-(1-p)^m} \qquad \square$$



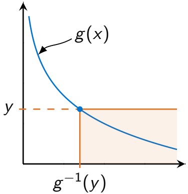


UNIVERSITY OF CAMBRIDGE
Department of Engineering

The CDF of $\mathrm{Y} = g(\mathrm{X})$ is $F_{\mathrm{Y}}(y) = \mathbb{P}[\mathrm{Y} \leq y] = \mathbb{P}[g(\mathrm{X}) \leq y]$



- If g is strictly increasing:
  $g(\mathrm{X}) \leq y \Leftrightarrow \mathrm{X} \leq g^{-1}(y)$
- $F_{\mathrm{Y}}(y) = \mathbb{P}[\mathrm{X} \leq g^{-1}(y)]$
  $= F_{\mathrm{X}}(g^{-1}(y))$
- Taking $\frac{\mathrm{d}}{\mathrm{d}y}$ on both sides,
  $$f_{\mathrm{Y}}(y) = \frac{f_{\mathrm{X}}\big(g^{-1}(y)\big)}{g'\big(g^{-1}(y)\big)}$$



- If g is strictly decreasing:
  $g(\mathrm{X}) \leq y \Leftrightarrow \mathrm{X} \geq g^{-1}(y)$
- $F_{\mathrm{Y}}(y) = \mathbb{P}[\mathrm{X} \geq g^{-1}(y)]$
  $= 1 - F_{\mathrm{X}}(g^{-1}(y))$
- Taking $\frac{\mathrm{d}}{\mathrm{d}y}$ on both sides,
  $$f_{\mathrm{Y}}(y) = -\frac{f_{\mathrm{X}}\big(g^{-1}(y)\big)}{g'\big(g^{-1}(y)\big)}$$

Hence, for g strictly monotonic, $f_{\mathrm{Y}=g(\mathrm{X})}(y) = \dfrac{f_{\mathrm{X}}\big(g^{-1}(y)\big)}{\left|g'\big(g^{-1}(y)\big)\right|}$

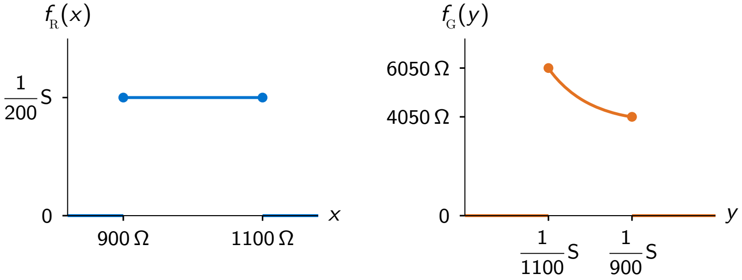
UNIVERSITY OF CAMBRIDGE
Department of Engineering

Suppose that a resistance R is uniformly distributed between 900 and 1100 Ω, what is the PDF of its conductance $\mathrm{G} = \dfrac{1}{\mathrm{R}}$?

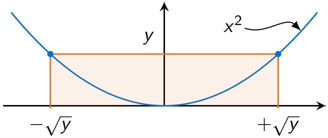
Here $g(x) = \frac{1}{x}$ is strictly decreasing, so we can use $f_{\mathrm{G}}(y) = \frac{f_{\mathrm{X}}(g^{-1}(y))}{|g'(g^{-1}(y))|}$ with

$$g'(x) = -\frac{1}{x^2} \quad \text{and} \quad g^{-1}(y) = \frac{1}{y} \quad \text{hence} \quad |g'(g^{-1}(y))| = y^2$$

$$P_{\mathrm{R}}(x) = \begin{cases} \frac{1}{200} & \text{if } x \in [900, 1100] \\ 0 & \text{otherwise} \end{cases} \Rightarrow P_{\mathrm{G}}(y) = \begin{cases} \frac{1}{200y^2} & \text{if } y \in [\frac{1}{1100}, \frac{1}{900}] \\ 0 & \text{otherwise} \end{cases} \qquad \square$$

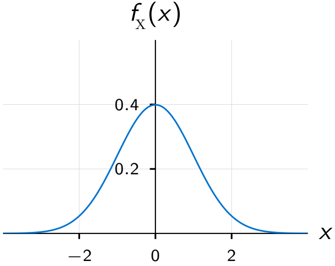


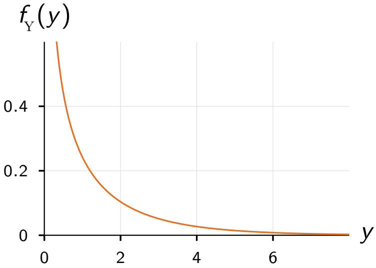
With $\mathrm{X} \sim \mathcal{N}(0,1)$, what is the PDF of $\mathrm{Y} = \mathrm{X}^2$?
Here $g(x)$ is not monotonic, and $\mathrm{Y} \geq 0$.



$F_{\mathrm{Y}}(y) = \mathbb{P}[\mathrm{Y} \leq y] = \mathbb{P}[-\sqrt{y} \leq \mathrm{X} \leq +\sqrt{y}] = F_{\mathrm{X}}(+\sqrt{y}) - F_{\mathrm{X}}(-\sqrt{y})$.

Taking $\frac{\mathrm{d}}{\mathrm{d}y}$ on both sides, $f_{\mathrm{Y}}(y) = \frac{f_{\mathrm{X}}(\sqrt{y})}{2\sqrt{y}} + \frac{f_{\mathrm{X}}(-\sqrt{y})}{2\sqrt{y}} = \frac{e^{-y/2}}{\sqrt{2\pi y}}$ for $y \geq 0$ $\square$





- What is the PDF of $\mathrm{Y} = a\mathrm{X} + b$?
  Here $g(x)$ is strictly monotonic and we can use the formula $f_{\mathrm{Y}}(y) = \frac{f_{\mathrm{X}}(g^{-1}(y))}{|g'(g^{-1}(y))|}$. Hence,

$$f_{\mathrm{Y}}(y) = \frac{1}{|a|} f_{\mathrm{X}}\left(\frac{y-b}{a}\right)$$

  Note that $\mathbb{E}[a\mathrm{X} + b] = a\mathbb{E}[\mathrm{X}] + b$ and $\mathrm{Var}[a\mathrm{X} + b] = a^2\mathrm{Var}[\mathrm{X}]$.
- With $\mathrm{X} \sim \mathcal{N}(\mu, \sigma^2)$, what is the PDF of $\mathrm{Y} = \frac{\mathrm{X}-\mu}{\sigma}$?
  With $a = 1/\sigma$ and $b = -\mu/\sigma$, we find

$$f_{\mathrm{Y}}(y) = \sigma f_{\mathrm{X}}(\sigma y + \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \quad \square$$

  so $\mathrm{Y} \sim \mathcal{N}(0, 1)$ standard Gaussian (as stated in the previous lecture).

UNIVERSITY OF CAMBRIDGE
Department of Engineering

Let X and Y be two random variables. We are interested in the probability distribution of the random variable $\mathrm{S} = \mathrm{X} + \mathrm{Y}$.

- We have already seen $\mathbb{E}[\mathrm{S}] = \mathbb{E}[\mathrm{X}] + \mathbb{E}[\mathrm{Y}]$.
- One can easily show that
$$\mathrm{Var}[\mathrm{S}] = \mathrm{Var}[\mathrm{X}] + \mathrm{Var}[\mathrm{Y}] + 2\mathrm{Cov}[\mathrm{X}, \mathrm{Y}]$$
where $\mathrm{Cov}[\mathrm{X}, \mathrm{Y}] = \mathbb{E}[\mathrm{XY}] - \mathbb{E}[\mathrm{X}]\mathbb{E}[\mathrm{Y}]$ is called the covariance.
- We also define the correlation coefficient:
$$\rho = \frac{\mathrm{Cov}[\mathrm{X}, \mathrm{Y}]}{\sqrt{\mathrm{Var}[\mathrm{X}]\mathrm{Var}[\mathrm{Y}]}}$$
which satisfies $-1 < \rho < 1$.
- If $\rho < 0$, X and Y are *anticorrelated*
If $\rho > 0$, X and Y are *correlated*
If $\rho = 0$, X and Y are *uncorrelated* (and $\mathbb{E}[\mathrm{XY}] = \mathbb{E}[\mathrm{X}]\mathbb{E}[\mathrm{Y}]$)

Let X and Y be two discrete random variables, with joint PMF $P_{\mathrm{XY}}(x,y)$, and $\mathrm{S}=\mathrm{X}+\mathrm{Y}$. Using the law of total probability

$$P_{\mathrm{S}}(s)=\sum_{y\in\mathbb{Y}}P_{\mathrm{S|Y}}(s|y)P_{\mathrm{Y}}(y)$$

with

$$\begin{aligned}P_{\mathrm{S|Y}}(s|y)=\mathbb{P}[\mathrm{S}=s|\mathrm{Y}=y]&=\mathbb{P}[\mathrm{X}+\mathrm{Y}=s|\mathrm{Y}=y]\\&=\mathbb{P}[\mathrm{X}=s-y|\mathrm{Y}=y]\\&=P_{\mathrm{X|Y}}(s-y|y)\end{aligned}$$

So

$$\begin{aligned}P_{\mathrm{S}}(s)=\sum_{y\in\mathbb{Y}}P_{\mathrm{X|Y}}(s-y|y)P_{\mathrm{Y}}(y)&=\sum_{y\in\mathbb{Y}}P_{\mathrm{XY}}(s-y,y)\\&=\sum_{x\in\mathbb{X}}P_{\mathrm{XY}}(x,s-x)\end{aligned}$$

If X and Y are independent,

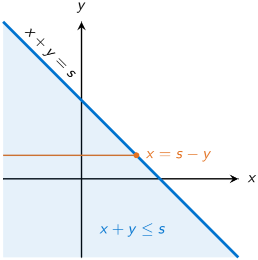
$$P_{\mathrm{S}}(s)=\sum_{y\in\mathbb{Y}}P_{\mathrm{X}}(s-y)P_{\mathrm{Y}}(y)=\sum_{x\in\mathbb{X}}P_{\mathrm{X}}(x)P_{\mathrm{Y}}(s-x)$$

that is, $P_{\mathrm{X+Y}}=P_{\mathrm{X}}*P_{\mathrm{Y}}$ the discrete convolution product

UNIVERSITY OF CAMBRIDGE
Department of Engineering

Let X and Y be two continuous random variables, with joint PDF $f_{\mathrm{XY}}(x,y)$, and $\mathrm{S} = \mathrm{X} + \mathrm{Y}$.

$$\begin{aligned} F_{\mathrm{S}}(s) &= \mathbb{P}[\mathrm{X} + \mathrm{Y} \leq s] \\ &= \iint_{x+y \leq s} f_{\mathrm{XY}}(x,y)\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{s-y} f_{\mathrm{XY}}(x,y)\mathrm{d}x\mathrm{d}y \\ &\overset{u=x+y}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{s} f_{\mathrm{XY}}(u-y,y)\mathrm{d}u\mathrm{d}y \end{aligned}$$



Take $\frac{\mathrm{d}}{\mathrm{d}s}$ on both sides

$$f_{\mathrm{S}}(s) = \int_{-\infty}^{\infty} f_{\mathrm{XY}}(s-y,y)\mathrm{d}y = \int_{-\infty}^{\infty} f_{\mathrm{XY}}(x,s-x)\mathrm{d}x$$

If X and Y are independent,

$$f_{\mathrm{S}}(s) = \int_{-\infty}^{\infty} f_{\mathrm{X}}(s-y)f_{\mathrm{Y}}(y)\mathrm{d}y = \int_{-\infty}^{\infty} f_{\mathrm{X}}(x)f_{\mathrm{Y}}(s-x)\mathrm{d}x$$

that is, $f_{\mathrm{X+Y}} = f_{\mathrm{X}} * f_{\mathrm{Y}}$ the convolution product

UNIVERSITY OF CAMBRIDGE
Department of Engineering

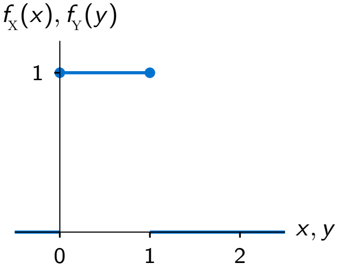
Let X and Y be independent and uniform between 0 and 1, that is

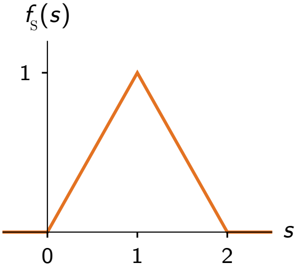
$$f_{\mathrm{X}}(x) = f_{\mathrm{Y}}(x) = \begin{cases} 1 & \text{if } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

and $\mathrm{S} = \mathrm{X} + \mathrm{Y}$. Find $f_{\mathrm{S}}(s)$.

$$f_{\mathrm{S}}(s) = \int_{-\infty}^{\infty} f_{\mathrm{X}}(s-y) f_{\mathrm{Y}}(y)\mathrm{d}y = \int_{0}^{1} f_{\mathrm{X}}(s-y)\mathrm{d}y \overset{u=s-y}{=} \int_{s-1}^{s} f_{\mathrm{X}}(u)\mathrm{d}u$$

$$= \begin{cases} \int_0^s \mathrm{d}u = s & \text{if } s \in [0,1] \\ \int_{s-1}^1 \mathrm{d}u = 2-s & \text{if } s \in [1,2] \\ 0 & \text{otherwise} \end{cases} \qquad \square$$





UNIVERSITY OF CAMBRIDGE
Department of Engineering

You have seen Fourier and Laplace Transforms in other courses.
You appreciate their usefulness in sidestepping tedious calculations.
Here we introduce two transforms with similar benefits:

- For a discrete random variable X with support $\mathbb{X}$, the Probability-Generating Function is defined by

$$G_{\mathrm{X}}(z) = \sum_{k \in \mathbb{X}} z^k P_{\mathrm{X}}(k) = \mathbb{E}[z^{\mathrm{X}}] \quad \text{[DB p.27]}$$

- For a continuous random variable X, the Moment-Generating Function is defined by

$$g_{\mathrm{X}}(s) = \int_{-\infty}^{+\infty} f_{\mathrm{X}}(x) e^{sx} \mathrm{d}x = \mathbb{E}[e^{s\mathrm{X}}] \quad \text{[DB p.28]}$$

- **Moments**
  Since $\frac{\mathrm{d}^k}{\mathrm{d}z^k}(z^{\mathrm{X}}) = \mathrm{X}(\mathrm{X}-1)(\mathrm{X}-2)\ldots(\mathrm{X}-k+1)z^{\mathrm{X}-k}$, we verify that

$$\mathbb{E}[\mathrm{X}] = G'_{\mathrm{X}}(1) \quad \text{[DB p.27]}$$
$$\mathrm{Var}[\mathrm{X}] = G''_{\mathrm{X}}(1) + G'_{\mathrm{X}}(1) - G'_{\mathrm{X}}(1)^2 \quad \text{[DB p.27]}$$

- **Sum of independent random variables**
  For X and Y two independent discrete random variables, $\mathbb{E}[z^{\mathrm{X}+\mathrm{Y}}] = \mathbb{E}[z^{\mathrm{X}} z^{\mathrm{Y}}] = \mathbb{E}[z^{\mathrm{X}}]\mathbb{E}[z^{\mathrm{Y}}]$ so

$$G_{\mathrm{X}+\mathrm{Y}}(z) = G_{\mathrm{X}}(z) \times G_{\mathrm{Y}}(z)$$

  More generally, for $\{\mathrm{X}_1, \mathrm{X}_2 \ldots \mathrm{X}_n\}$ mutually independent,

$$G_{\sum_{i=1}^n \mathrm{X}_i}(z) = \prod_{i=1}^{n} G_{\mathrm{X}_i}(z)$$

UNIVERSITY OF CAMBRIDGE
Department of Engineering

With $p \in [0,1]$, $q = 1-p$, $n \in \{1,2,\dots\}$ and $\lambda > 0$:

Distribution of X	Support of X	$P_X(k)$	$G_X(z)$
$\mathrm{Ber}(p)$	$\{0,1\}$	$p^k q^{1-k}$	$q + pz$
$\mathrm{Geo}(p)$	$\{1,2\dots\}$	pq^{k-1}	$\dfrac{pz}{1-qz}$ $(\lvert z\rvert < q^{-1})$
$\mathrm{B}(n,p)$	$\{0,1\dots,n\}$	${}^n\mathrm{C}_k\, p^k q^{n-k}$	$(q+pz)^n$
$\mathrm{Pois}(\lambda)$	$\{0,1\dots\}$	$\dfrac{\lambda^k e^{-\lambda}}{k!}$	$e^{\lambda(z-1)}$

- Note the short form for the Bernoulli PMF...
- Using the expressions of G_X and $G_{X+Y} = G_X \times G_Y$, one gets

same p

$$X \sim \mathrm{B}(n_X, p), \quad Y \sim \mathrm{B}(n_Y, p) \quad \Rightarrow \quad X+Y \sim \mathrm{B}(n_X+n_Y, p)$$

$$X \sim \mathrm{Pois}(\lambda_X), \quad Y \sim \mathrm{Pois}(\lambda_Y) \quad \Rightarrow \quad X+Y \sim \mathrm{Pois}(\lambda_X+\lambda_Y)$$

when X and Y are independent.

UNIVERSITY OF CAMBRIDGE
Department of Engineering

- $\mathrm{X} \sim \mathrm{Ber}(p)$

  Immediate from definition $\displaystyle\sum_{k=0}^{1} z^k P_{\mathrm{X}}(k) = (1-p)z^0 + pz^1$ $\square$

- $\mathrm{X} \sim \mathrm{Geo}(p)$

  From definition

  $$\sum_{k=1}^{\infty} pq^{k-1} z^k = pz \sum_{k=1}^{\infty} (qz)^{k-1} \overset{j=k-1}{=} pz \sum_{j=0}^{\infty} (qz)^j \overset{|qz|<1}{=} \frac{pz}{1-qz} \qquad \square$$

- $\mathrm{X} \sim \mathrm{B}(n, p)$

  $$\sum_{k=0}^{n} {}^n\mathrm{C}_k \, p^k q^{n-k} z^k = \sum_{k=0}^{n} {}^n\mathrm{C}_k \, (zp)^k q^{n-k} \overset{\text{Binomial expansion}}{=} (q + pz)^n \qquad \square$$

- $\mathrm{X} \sim \mathrm{Pois}(\lambda)$

  $$\sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} z^k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda z)^k}{k!} \overset{\text{Power series of } e^{\lambda z}}{=} e^{-\lambda} e^{\lambda z} \qquad \square$$

UNIVERSITY OF CAMBRIDGE
Department of Engineering

- **Moments**
  Since $\frac{\mathrm{d}^k}{\mathrm{d}s^k}(e^{sX}) = X^k e^{sX}$, we verify that $g_X^{(k)}(0) = \mathbb{E}[X^k]$. In particular:
  $$\mathbb{E}[X] = g_X'(0) \quad \text{[DB p.28]}$$
  $$\mathrm{Var}[X] = g_X''(0) - g_X'(0)^2 \quad \text{[DB p.28]}$$

- **Sum of independent random variables**
  For X and Y two independent continuous random variables, $\mathbb{E}[e^{s(X+Y)}] = \mathbb{E}[e^{sX}e^{sY}] = \mathbb{E}[e^{sX}]\mathbb{E}[e^{sY}]$ so
  $$g_{X+Y}(s) = g_X(s) \times g_Y(s)$$
  More generally, for $\{X_1, X_2 \ldots X_n\}$ mutually independent,
  $$g_{\sum_{i=1}^n X_i}(s) = \prod_{i=1}^n g_{X_i}(s)$$

UNIVERSITY OF CAMBRIDGE
Department of Engineering

With $a < b \in \mathbb{R}$, $\lambda > 0$, $\mu \in \mathbb{R}$, $\sigma > 0$, $\alpha, \beta > 0$ and $\gamma = \alpha + \beta$:

Distribution	$\mathbb{X}$	$f_{\mathrm{X}}(x)$	$g_{\mathrm{X}}(s)$
$\mathrm{U}(a, b)$	$[a, b]$	$\dfrac{1}{b-a}$	$\dfrac{e^{bs} - e^{as}}{s(b-a)}$
$\mathrm{Exp}(\lambda)$	$\mathbb{R}^+$	$\lambda e^{-\lambda x}$	$\dfrac{\lambda}{\lambda - s}$ $(s<\lambda)$
$\mathcal{N}(\mu, \sigma^2)$	$\mathbb{R}$	$\dfrac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$	$e^{\mu s + \frac{\sigma^2 s^2}{2}}$
$\mathrm{Beta}(\alpha, \beta)$	$[0, 1]$	$\dfrac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$1 + \sum_{k=1}^{\infty} \left[\prod_{r=0}^{k-1} \dfrac{\alpha+r}{\gamma+r}\right] \dfrac{s^k}{k!}$

- Using the expressions of g_{X} and $g_{\mathrm{X+Y}} = g_{\mathrm{X}} \times g_{\mathrm{Y}}$, one gets
$$\mathrm{X} \sim \mathcal{N}(\mu_{\mathrm{X}}, \sigma_{\mathrm{X}}^2),\ \mathrm{Y} \sim \mathcal{N}(\mu_{\mathrm{Y}}, \sigma_{\mathrm{Y}}^2) \Rightarrow \mathrm{X+Y} \sim \mathcal{N}(\mu_{\mathrm{X}} + \mu_{\mathrm{Y}}, \sigma_{\mathrm{X}}^2 + \sigma_{\mathrm{Y}}^2)$$
when X and Y are independent.

UNIVERSITY OF CAMBRIDGE
Department of Engineering

- $\mathrm{X} \sim \mathrm{U}(a, b)$

  From definition $g_{\mathrm{X}}(s) = \frac{1}{b-a}\int_a^b e^{sx}\mathrm{d}x = \frac{1}{b-a}\left[\frac{e^{sx}}{s}\right]_a^b$ □

- $\mathrm{X} \sim \mathrm{Exp}(\lambda)$

  From definition

  $g_{\mathrm{X}}(s) = \lambda\int_0^\infty e^{(s-\lambda)x}\mathrm{d}x = \lambda\left[\frac{e^{(s-\lambda)x}}{s-\lambda}\right]_0^\infty \overset{s<\lambda}{=} \frac{\lambda}{\lambda - s}$ □

- $\mathrm{X} \sim \mathcal{N}(\mu, \sigma^2)$

  For the standard Gaussian, $g_{\mathrm{X}}(s) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}e^{sx}\mathrm{d}x =$

  $e^{s^2/2}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(x-s)^2/2}\mathrm{d}x = e^{s^2/2}$ and note that

  $g_{\sigma\mathrm{X}+\mu}(s) = \mathbb{E}[e^{(\sigma s)\mathrm{X}+\mu s}] = e^{\mu s}g_{\mathrm{X}}(\sigma s)$ □

- $\mathrm{X} \sim \mathrm{Beta}(\alpha, \beta)$

  Tedious, we won't do it here. □

Let X_1, X_2 ... be independent random variables with means μ_1, $\mu_2 \ldots$ and variances σ_1^2, σ_2^2 ... Then the random variable

$$\mathrm{S}_n = \mathrm{X}_1 + \mathrm{X}_2 + \cdots + \mathrm{X}_n$$

tends to a Gaussian random variable S,

$$\mathrm{S} \sim \mathcal{N}(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)$$

as n tends to infinity, regardless of the actual individual distributions of X_i.

- As expected from independence, means and variances add up.
- If all X_i are *identically* distributed with mean μ and variance σ^2, then the theorem is equivalent to

$$\mathrm{Y}_n = \frac{\mathrm{X}_1 + \mathrm{X}_2 + \cdots + \mathrm{X}_n - n\mu}{\sigma\sqrt{n}} \quad \overset{n\to\infty}{\sim} \quad \mathcal{N}(0, 1)$$

Let us consider X_1, $X_2 \dots X_n$ be independent random variables that have been *shifted/rescaled* to have means $\mu_1 = \mu_2 = \cdots = \mu_n = 0$ and variances $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_2^2 = 1$.

We will show that $Y_n = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}$ tends to a standard Gaussian random variable when $n \to \infty$.

We write the MGF of Y_n as

$$g_{Y_n}(s) = \prod_{k=1}^{n} g_{X_k}\left(\tfrac{s}{\sqrt{n}}\right) \overset{\text{Taylor series}}{=} \prod_{k=1}^{n} \left[g_{X_k}(0) + g'_{X_k}(0)\tfrac{s}{\sqrt{n}} + g''_{X_k}(0)\tfrac{s^2}{2n} + o\left(\tfrac{s^3}{n^{3/2}}\right) \right]$$

Since $g_X^{(k)}(0) = \mathbb{E}[X^k]$, $g_{X_k}(0) = 1$, $g'_{X_k}(0) = \mu_k = 0$ and $g''_{X_k}(0) = \sigma_k^2 = 1$ and we are left with

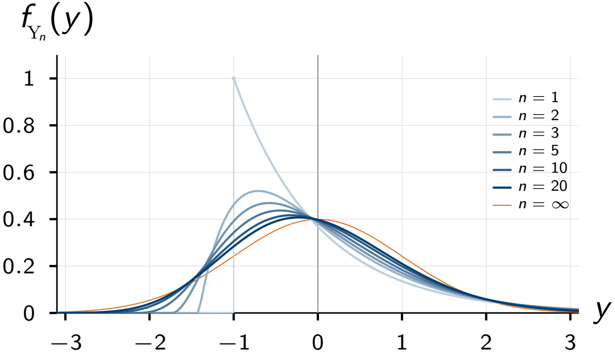
$$g_{Y_n}(s) = \prod_{k=1}^{n} \left[1 + \tfrac{s^2}{2n} + o\left(\tfrac{s^3}{n^{3/2}}\right) \right] = \left(1 + \tfrac{s^2}{2n}\right)^n + o\left(\tfrac{s^3}{n^{3/2}}\right)$$

and $\lim_{n\to\infty} g_{Y_n}(s) = \lim_{n\to\infty} \left(1 + \frac{s^2}{2n}\right)^n = e^{s^2/2}$ which is the MGF of the standard Gaussian. Hence $Y_\infty \sim \mathcal{N}(0, 1)$. $\square$

There are subtle restrictions on the distribution of X_k for the theorem to work. But it works in most cases. . .

Let us consider that $\mathrm{X}_i \sim \mathrm{Exp}(1)$ for all $i \in \{1, \ldots, n\}$
(ie., $\mu_i = \mathbb{E}[\mathrm{X}_i] = 1 = \mu$ and $\sigma_i^2 = \mathrm{Var}[\mathrm{X}_i] = 1 = \sigma^2$ for all i).

Let us monitor the PDF of $\mathrm{Y}_n = \dfrac{\mathrm{X}_1 + \mathrm{X}_2 + \cdots + \mathrm{X}_n - n\mu}{\sigma\sqrt{n}}$ as n increases. (Note that $\mathrm{Y}_1 = \mathrm{X}_1 - 1$ is shifted exponential).



- We haven't seen any example of a *joint* PDF;
- We know that the Gaussian distribution is important;
- We can concisely write the joint PDF $f_{\mathrm{X}_1\mathrm{X}_2\ldots\mathrm{X}_n}(x_1, x_2 \ldots, x_n)$ of n random variables as $f_{\mathbf{X}}(\boldsymbol{x})$ using the vector notation.

Hence, the n-dimensional random vector $\mathbf{X} = [\mathrm{X}_1, \mathrm{X}_2 \ldots, \mathrm{X}_n]^\mathsf{T}$ is multivariate Gaussian if

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \;\Leftrightarrow\; f_{\mathbf{X}}(\boldsymbol{x}) = \frac{\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\mathsf{T}\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{x}-\boldsymbol{\mu})\right)}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}}$$

evaluated at $\boldsymbol{x} = [x_1, x_2 \ldots, x_n]^\mathsf{T}$, with

- $\boldsymbol{\mu} = [\mu_1, \mu_2 \ldots, \mu_n]^\mathsf{T}$ is the mean vector:
  $\mu_i = \mathbb{E}[\mathrm{X}_i]$ for $i = 1 \ldots n$
- $\boldsymbol{\Sigma}$ is the symmetric $n \times n$ covariance matrix:
  $$\Sigma_{ij} = \mathbb{E}[(\mathrm{X}_i - \mu_i)(\mathrm{X}_j - \mu_j)] = \begin{cases} \mathrm{Cov}[\mathrm{X}_i, \mathrm{X}_j] & \text{if } i \neq j \\ \mathrm{Var}[\mathrm{X}_i] = \sigma_i^2 & \text{if } i = j \end{cases}$$

For $n = 2$, we have $\begin{bmatrix} \mathrm{X}_1 \\ \mathrm{X}_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right)]$ where $\rho = \frac{\mathrm{Cov}[\mathrm{X}_1,\mathrm{X}_2]}{\sigma_1\sigma_2}$ is the correlation. The full expression

$$f_{\mathrm{X}_1\mathrm{X}_2}(x_1, x_2) = \frac{\exp\left(-\frac{1}{2(1-\rho^2)} \left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 \right] \right)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

In particular, we can show:

- The marginals are Gaussian: $f_{\mathrm{X}_1} = \mathcal{N}(\mu_1, \sigma_1^2)$, $f_{\mathrm{X}_2} = \mathcal{N}(\mu_2, \sigma_2^2)$.
  This is also true for any "partial" marginals (integration over $k < n$ components of $\mathbf{X}$) of $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- If $\rho = 0$, then $f_{\mathrm{X}_1\mathrm{X}_2} = f_{\mathrm{X}_1} \times f_{\mathrm{X}_2}$ and $\mathrm{X}_1, \mathrm{X}_2$ are independent.
  The components of $\mathbf{X}$ are mutually independent if $\boldsymbol{\Sigma}$ is diagonal.
- The conditional $f_{\mathrm{X}_1|\mathrm{X}_2 = x_2} = \mathcal{N}\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), (1-\rho^2)\sigma_1^2\right)$ is also Gaussian.
  The multivariate conditionals $f_{\mathrm{X}_1 \ldots \mathrm{X}_k|\mathrm{X}_{k+1} \ldots \mathrm{X}_n}$ are also Gaussian.

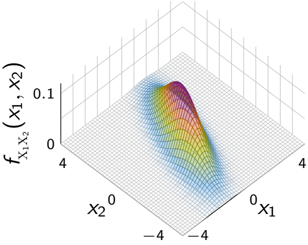
UNIVERSITY OF CAMBRIDGE
Department of Engineering

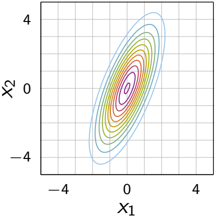
$\rho = 0.7$

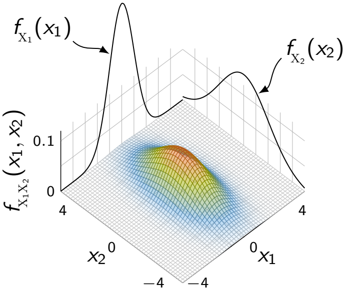
$\mu_1 = 0$
$\mu_2 = 0$

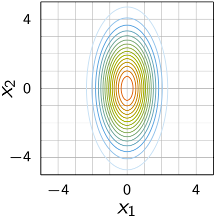
$\sigma_1^2 = 1$
$\sigma_2^2 = 4$

$\rho = 0$









You can attempt Problems 1 to 8 of Examples Paper 6