EE2S31 Signal Processing – Stochastic Processes

Lecture 3: Sums of RVs & The Sample Mean – Chs. 6, 9 & 10

Alle-Jan van der Veen

–

2 May 2022

Today

- Given random variables X and Y. What is the PDF of $W = X + Y$?
  - Transformed RVs $\Rightarrow$ for iid RVs, **convolution of PDFs**.
  - Easier: Using **moment generating functions** (Laplace transform of PDF)
- **Expected value and sample mean**
  - Expected value: $\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x) \mathrm{d}x$.
  - What if $f_X(x)$ is unknown? $\Rightarrow$ use sample mean of X:

$$M_n(X) = \frac{X_1 + \cdots + X_n}{n}.$$

  - How good is $M_n(X)$ as an approximation of $\mathrm{E}[X]$?

(Ch. 6.2) Derived random variables – continuous RVs

How can we compute the PDF of derived RVs $\boldsymbol{Y} = g(\boldsymbol{X})$:

- **Special (simple) case:** for linear transformations, we saw
  - For scalars: $Y = aX + b \;\Leftrightarrow\; f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$
  - For vectors: $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b} \;\Leftrightarrow\; f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{|\det(\boldsymbol{A})|} f_{\boldsymbol{X}}\left(\boldsymbol{A}^{-1}(\boldsymbol{y} - \boldsymbol{b})\right)$

- **General approach,** using CDFs:
  (1) Find the CDF $F_{\boldsymbol{X}}(\boldsymbol{x}) = \mathrm{P}[\boldsymbol{X} \leq \boldsymbol{x}]$
  (2) Transform to $F_{\boldsymbol{Y}}(\boldsymbol{y}) = \mathrm{P}[g(\boldsymbol{X}) \leq \boldsymbol{y}] \stackrel{?}{=} \mathrm{P}[\boldsymbol{X} \leq g^{-1}(\boldsymbol{y})]$
  This requires $g^{-1}(\boldsymbol{y})$ and a check on "$\leq$"
  (3) Compute the PDF by calculating $f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{\mathrm{d}F_{\boldsymbol{Y}}(\boldsymbol{y})}{\mathrm{d}\boldsymbol{y}}$

Problem 6.2.2

X is a Gaussian(0,1) random variable. Find the CDF of $Y = |X|$, and its expected value $\mathrm{E}[Y]$.

Since $Y \geq 0$, $F_Y(y) = 0$ for $y < 0$. For $y \geq 0$,

$$F_Y(y) = \mathrm{P}[\,|X| \leq y\,] = \mathrm{P}[-y \leq X \leq y] = \Phi(y) - \Phi(-y) = 2\Phi(y) - 1$$

$$\frac{\mathrm{d}F_Y(y)}{\mathrm{d}y} = 2f_X(y) = \frac{2}{\sqrt{2\pi}} e^{-y^2/2}$$

Thus, the complete expression is

$$f_Y(y) = \begin{cases} \frac{2}{\sqrt{2\pi}} e^{-y^2/2} & y \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathrm{E}[Y] = \int_{-\infty}^{\infty} y\, f_Y(y) dy = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} y e^{-y^2/2} dy = -\sqrt{\frac{2}{\pi}} e^{-y^2/2} \Big|_0^{\infty} = \sqrt{\frac{2}{\pi}}$$

(Ch.6.5) PDF of the sum of two random variables

Special case: $W = X + Y$

$$f_W(w) = \int_{-\infty}^{\infty} f_{X,Y}(x, w-x)\mathrm{d}x = \int_{-\infty}^{\infty} f_{X,Y}(w-y, y)\mathrm{d}y$$

Proof:

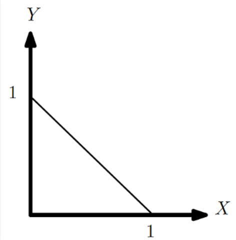
$$F_W(w) = \mathrm{P}[X + Y \leq w] \quad = \quad \int_{-\infty}^{\infty} \left(\int_{-\infty}^{w-x} f_{X,Y}(x,y)\mathrm{d}y \right) \mathrm{d}x$$

$$\begin{aligned} f_W(w) = \frac{\mathrm{d}F_W(w)}{\mathrm{d}w} \quad &= \quad \int_{-\infty}^{\infty} \left(\frac{\mathrm{d}}{\mathrm{d}w} \left(\int_{-\infty}^{w-x} f_{X,Y}(x,y)\mathrm{d}y \right) \right) \mathrm{d}x \\ &= \quad \int_{-\infty}^{\infty} f_{X,Y}(x, w-x)\mathrm{d}x \end{aligned}$$

Problem 6.5.2

X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 2 & x \geq 0, y \geq 0, x+y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$



Find the PDF of $W = X + Y$.

Write $f_W(w) = \int_{-\infty}^{\infty} f_{X,Y}(x, w-x)\mathrm{d}x$.

For $0 \leq w \leq 1$, $f_W(w) = \int_0^w 2\,\mathrm{d}x = 2w$.

For $w < 0$ or $w > 1$, $f_W(w) = 0$ since $0 \leq W \leq 1$. The complete expression is

$$f_W(w) = \begin{cases} 2w & 0 \leq w \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Sum of two independent random variables

For independent RVs: $f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$.

So, for two independent RVs X and Y we get

$$\begin{aligned} f_W(w) &= \int_{-\infty}^{\infty} f_{X,Y}(x, w-x)\mathrm{d}x \\ &= \int_{-\infty}^{\infty} f_X(x) f_Y(w-x)\mathrm{d}x \end{aligned}$$

- The PDF of the sum of two independent RVs is the convolution of the two PDFs. (Equivalent for discrete RVs.)

Problem 6.5.5

Random variables X and Y are independent exponential random variables with expected values $\mathrm{E}[X] = 1/\lambda$ and $\mathrm{E}[Y] = 1/\mu$.

If $\mu \neq \lambda$, what is the PDF of $W = X + Y$?

Problem 6.5.5

Random variables X and Y are independent exponential random variables with expected values $\mathrm{E}[X] = 1/\lambda$ and $\mathrm{E}[Y] = 1/\mu$.

If $\mu \neq \lambda$, what is the PDF of $W = X + Y$?

$W = X + Y$. Work out the convolution: (for $\lambda \neq \mu$, $x \geq 0, y \geq 0$)

$$
\begin{aligned}
f_X(w) &= \int_{-\infty}^{\infty} f_X(x) f_Y(w-x) \mathrm{d}x \\
&= \int_0^w \lambda e^{-\lambda x} \mu e^{-\mu(w-x)} \mathrm{d}x \qquad \text{since } y = w - x \geq 0 \Rightarrow x \leq w \\
&= \lambda\mu\, e^{-\mu w} \int_0^w e^{-(\lambda-\mu)x} \mathrm{d}x \\
&= \begin{cases} \dfrac{\lambda\mu}{\lambda-\mu}\left(e^{-\mu w} - e^{-\lambda w}\right) & w \geq 0 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

Expected value of sums of random variables

Consider the sum $W = X_1 + X_2 + \cdots + X_n$.

- The expected value $\mathrm{E}[W]$ is given by

$$\mathrm{E}[W] = \mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_n]$$

- The variance of W is given by

$$\mathrm{var}[W] = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{cov}[X_i, X_j] = \sum_{i=1}^{n} \mathrm{var}[X_i] + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathrm{cov}[X_i, X_j]$$

For uncorrelated variables we obtain $\mathrm{var}[W] = \sum_{i=1}^{n} \mathrm{var}[X_i]$

(Ch. 9) PDF of the sum of independent random variables

What about the PDF of the sum of more independent variables?

For $W = X + Y + Z$ (independent RVs)

$$f_W(w) = f_X(x) * f_Y(y) * f_Z(z)$$

- Calculating such convolutions is easier in frequency (or Laplace) domain.
- The Laplace transform of a PDF or MDF is called the **Moment Generating Function** (MGF).

Moment generating function

The moment generating function (MGF) is defined as the Laplace transform of the PDF:

$$\phi_X(s) := \int_{-\infty}^{\infty} f_X(x)\, e^{sx} \mathrm{d}x = \mathrm{E}[e^{sX}]\,, \qquad (s \in \mathrm{ROC})$$

- Note the missing "$-$" sign on s: different convention than in S&S. Also, s is limited to real values.
- Nonetheless, the usual properties of Laplace transforms apply:

$$f_W(w) = f_X(x) * f_Y(y) \quad \Leftrightarrow \quad \phi_W(s) = \phi_X(s) \cdot \phi_Y(s)$$

For discrete RVs, this looks like a z-transform of the PMF (with $z = e^s$)

Moment generating function: Properties

For continuous RVs:

$$\phi_X(s) = \int_{-\infty}^{\infty} f_X(x)\, e^{sx} \mathrm{d}x = \mathrm{E}[e^{sX}].$$

For discrete RVs:

$$\phi_X(s) = \sum_{x_i \in S_x} P_X(x_i)\, e^{sx_i} = \mathrm{E}[e^{sX}].$$

- $\phi_X(0) = \mathrm{E}[e^0] = 1$
- $\dfrac{\mathrm{d}\phi_X(s)}{\mathrm{d}s} = \int_{-\infty}^{\infty} x\, f_X(x) e^{sx} \mathrm{d}x \quad \Rightarrow \quad \left.\dfrac{\mathrm{d}\phi_X(s)}{\mathrm{d}s}\right|_{s=0} = \mathrm{E}[X]$
- $\left.\dfrac{\mathrm{d}^n\phi_X(s)}{\mathrm{d}s^n}\right|_{s=0} = \mathrm{E}[X^n]$

Example (1)

Let X be exponentially distributed (e.g., duration of a phone call):

$$f_X(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases}$$

What is the MGF $\phi_X(s)$?

Example (1)

Let X be exponentially distributed (e.g., duration of a phone call):

$$f_X(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases}$$

What is the MGF $\phi_X(s)$?

$$\begin{aligned} \phi_X(s) &= \mathrm{E}[e^{sx}] = \int_{-\infty}^{\infty} e^{sx} f_X(x) \mathrm{d}x = \int_0^{\infty} e^{sx} \lambda e^{-\lambda x} \mathrm{d}x \\ &= \int_0^{\infty} \lambda e^{(s-\lambda)x} \mathrm{d}x = \frac{\lambda}{s-\lambda} e^{(s-\lambda)x} \Big|_0^{\infty} \end{aligned}$$

Notice that integral only converges for $s - \lambda \leq 0$ (as $x \geq 0$).

The MGF is: $\phi_X(s) = \dfrac{\lambda}{\lambda - s}$ (ROC: $s < \lambda$)

Example (2)

Let X be exponentially distributed. Calculating

$$\mathrm{E}[X^n] = \int_0^\infty x^n f_X(x)\mathrm{d}x = \int_0^\infty x^n \lambda e^{-\lambda x}\mathrm{d}x$$

requires n times partial integration!

The MGF of X is $\phi_X(s) = \frac{\lambda}{\lambda - s}$, for $s < \lambda$

- $\mathrm{E}[X] = \frac{\mathrm{d}\phi_X(s)}{\mathrm{d}s}\Big|_{s=0} = \frac{\lambda}{(\lambda - s)^2}\Big|_{s=0} = \frac{1}{\lambda}$
- $\mathrm{E}[X^2] = \frac{\mathrm{d}^2\phi_X(s)}{\mathrm{d}s^2}\Big|_{s=0} = \frac{2\lambda}{(\lambda - s)^3}\Big|_{s=0} = \frac{2}{\lambda^2}$
- $\mathrm{E}[X^n] = \frac{\mathrm{d}^n\phi_X(s)}{\mathrm{d}s^n}\Big|_{s=0} = \frac{n!\lambda}{(\lambda - s)^{n+1}}\Big|_{s=0} = \frac{n!}{\lambda^n}$

Using MGFs, we only need to calculate n derivatives for $\mathrm{E}[X^n]$.

Problem 9.2.1

For a constant $a > 0$, a Laplace random variable X has PDF

$$f_X(x) = \frac{a}{2}e^{-a|x|}, \quad -\infty < x < \infty$$

Calculate the MGF $\phi_X(s)$.

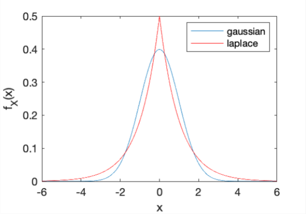
Problem 9.2.1

For a constant $a > 0$, a Laplace random variable X has PDF

$$f_X(x) = \frac{a}{2} e^{-a|x|}, \quad -\infty < x < \infty$$

Calculate the MGF $\phi_X(s)$.

$$\begin{aligned} \phi_X(s) &= E[e^{sX}] = \frac{a}{2}\int_{-\infty}^{0} e^{sx} e^{ax} dx + \frac{a}{2}\int_{0}^{\infty} e^{sx} e^{-ax} dx \\ &= \frac{a}{2} \frac{e^{(s+a)x}}{s+a}\bigg|_{-\infty}^{0} + \frac{a}{2} \frac{e^{(s-a)x}}{s-a}\bigg|_{0}^{\infty} \\ &= \frac{a^2}{a^2 - s^2} \end{aligned}$$



Check ROC: $\{s + a \geq 0\} \cap \{s - a \leq 0\} = \{-a \leq s \leq a\}$.

- The Laplace distribution has "fat tails" and is often used to model noise that also has outliers

Problem 9.2.2

Random variables J and K have the joint probability mass function

$P_{J,K}(j,k$	$k=-1$	$k=0$	$k=1$	Total
$j=-2$	0.42	0.12	0.06	0.6
$j=-1$	0.28	0.08	0.04	0.4
Total	0.7	0.2	0.1	

Note: J and K are independent

(a) What is the MGF of J?

(b) What is the MGF of K?

(c) Find the PMF of $M = J + K$

(d) What is $\mathrm{E}[M^4]$?

Problem 9.2.2

Random variables J and K have the joint probability mass function

$P_{J,K}(j,k$	$k=-1$	$k=0$	$k=1$	Total
$j=-2$	0.42	0.12	0.06	0.6
$j=-1$	0.28	0.08	0.04	0.4
Total	0.7	0.2	0.1	

Note: J and K are independent

(a) What is the MGF of J?

(b) What is the MGF of K?

(c) Find the PMF of $M = J + K$

(d) What is $\mathrm{E}[M^4]$?

$$\phi_J(s) = 0.6e^{-2s} + 0.4e^{-s}$$
$$\phi_K(s) = 0.7e^{-s} + 0.2 + 0.1e^{s}$$

$$\begin{aligned}\phi_M(s) &= \phi_J(s) \cdot \phi_K(s) \\ &= 0.42e^{-3s} + (0.28 + 0.12)e^{-s} \\ &\quad + (0.06 + 0.08)e^{-s} + 0.04 \\ &= 0.42e^{-3s} + 0.4s^{-2s} \\ &\quad + 0.14e^{-s} + 0.04\end{aligned}$$

$$P_M(m) = \begin{cases} 0.42 & m = -3 \\ 0.40 & m = -2 \\ 0.14 & m = -1 \\ 0.04 & m = 0 \\ 0 & \text{otherwise} \end{cases}$$

Problem 9.2.2 (cont'd)

$$\phi_M(s) = 0.42e^{-3s} + 0.4s^{-2s} + 0.14e^{-s} + 0.04$$

$$\begin{aligned}
\frac{\mathrm{d}^4\phi_M(s)}{\mathrm{d}s^4} &= (-3)^4 0.42e^{-3s} + (-2)^4 0.4e^{-2s} + (-1)^4 0.14e^{-s} \\
\mathrm{E}[M^4] &= \left.\frac{\mathrm{d}^4\phi_M(s)}{\mathrm{d}s^4}\right|_{s=0} \\
&= (-3)^4 0.42 + (-2)^4 0.4 + (-1)^4 0.14 = 40.434
\end{aligned}$$

Compare to a direct calculation:

$$\begin{aligned}
\mathrm{E}[M^4] &= \sum_m P_M(m) m^4 \\
&= 0.42(-3)^4 + 0.4(-2)^4 + 0.14(-1)^4 + 0.04(0)^4 = 40.434
\end{aligned}$$

MGFs of standard distributions (Table 9.1/Appendix A)

Discrete RVs:

- Bernoulli(p):

$$P_X(x) = \begin{cases} 1-p & x = 0 \\ p & x = 1 \\ 0 & \text{otherwise} \end{cases} \quad \Leftrightarrow \quad \phi_X(s) = 1 - p + pe^s$$

- Binomial(n, p):

$$P_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \Leftrightarrow \quad \phi_X(s) = (1 - p + pe^s)^n$$

- Uniform($0, N-1$):

$$P_X(x) = \begin{cases} \frac{1}{N} & x = 0, \cdots, N-1 \\ 0 & \text{otherwise} \end{cases} \quad \Leftrightarrow \quad \phi_X(s) = \frac{1}{N} \frac{1 - e^{sN}}{1 - e^s}$$

MGFs of standard distributions (Table 9.1/Appendix A)

Continuous RVs:

- Gaussian(μ, σ):

$$f_X(x) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}} \qquad \Leftrightarrow \qquad \phi_X(s) = e^{s\mu+\sigma^2 s^2/2}$$

- Exponential(λ):

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases} \qquad \Leftrightarrow \qquad \phi_X(x) = \frac{\lambda}{\lambda - s}$$

- Laplace(a):

$$f_X(x) = \frac{a}{2} e^{-a|x|} \qquad \Leftrightarrow \qquad \phi_X(x) = \frac{a^2}{a^2 - s^2}$$

Problem 9.2.4

Let X be a Gaussian$(0, \sigma)$ random variable. Use the moment generating function to show that

$$\begin{aligned} E[X] &= 0\,, \qquad & E[X^2] &= \sigma^2 \\ E[X^3] &= 0\,, \qquad & E[X^4] &= 3\sigma^4 \end{aligned}$$

Use Appendix A: $\phi_X(s) = e^{\sigma^2 s^2/2}$

$$\begin{aligned} E[X] &= \sigma^2 s\, e^{\sigma^2 s^2/2}\big|_{s=0} &= 0 \\ E[X^2] &= \sigma^2 e^{\sigma^2 s^2/2} + \sigma^4 s^2 e^{\sigma^2 s^2/2}\big|_{s=0} &= \sigma^2 \\ E[X^3] &= (3\sigma^4 s + \sigma^6 s^3) e^{\sigma^2 s^2/2}\big|_{s=0} &= 0 \\ E[X^4] &= (3\sigma^4 + 6\sigma^6 s^2 + \sigma^8 s^4) e^{\sigma^2 s^2/2}\big|_{s=0} &= 3\sigma^4 \end{aligned}$$

MGF of linearly transformed RVs

The MGF of $Y = aX + b$ is $\phi_Y(s) = \mathrm{E}[e^{s(aX+b)}] = e^{sb}\phi_X(as)$

The MGF for sums of RVs

The MGF of a sum of n independent RVs

$$W = X_1 + \cdots + X_n$$

is given by

$$\phi_W(s) = \mathrm{E}[e^{sW}] = \mathrm{E}\left[e^{s\sum_{i=1}^{n} X_i}\right] = \mathrm{E}\left[\prod_{i=1}^{n} e^{sX_i}\right] = \prod_{i=1}^{n} \phi_{X_i}(s)$$

The sum of Gaussian RVs

Let $X_1, X_2, \cdots, X_n$ denote a sequence of independent Gaussian RVs.

What is the distribution of $W = X_1 + X_2 + \cdots + X_n$?

$$\begin{aligned}\phi_W(s) &= \phi_{X_1}(s)\,\phi_{X_2}(s)\ldots\phi_{X_n}(s)\\ &= e^{s\mu_1+\sigma_1^2 s^2/2}\, e^{s\mu_2+\sigma_2^2 s^2/2}\ldots e^{s\mu_n+\sigma_n^2 s^2/2}\\ &= e^{s(\mu_1+\mu_2+\cdots+\mu_n)+(\sigma_1^2+\sigma_2^2+\cdots+\sigma_n^2)s^2/2}\end{aligned}$$

- The distribution of a sum of independent Gaussians is again Gaussian with mean $\mu_1 + \mu_2 + \cdots + \mu_n$ and variance $\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$

The Central Limit Theorem

Given a sequence of iid random variables $X_1, X_2, \ldots, X_n$, each with expected value μ_X and variance σ_X^2.

Consider the *standardized sum* (i.e., normalized to mean 0, std 1):

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu_X}{\sqrt{n\sigma_X^2}}$$

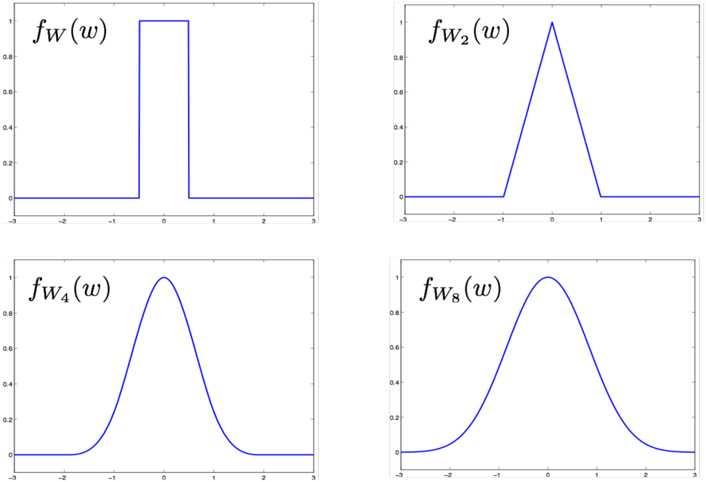
The CDF of Z_n then has the property:

$$\lim_{n\to\infty} F_{Z_n}(z) = \Phi(z).$$

- This means: if n becomes "large", the distribution of the sum of iid random variables approaches a Gaussian distribution.
- In practice, n does not have to be very large

The Central Limit Theorem: illustration

$W_n = \sum_n X_i$, with X_i a Uniform$(-\frac{1}{2}, \frac{1}{2})$ distribution



Problem 9.4.9 — Use of CLT

Let X_i be Uniform(-1,1). Let $Y_i = 20 + 15X_i^2$. Let $W = \frac{1}{100}\sum_{i=1}^{100} Y_i$.

Estimate $\mathrm{P}[W \leq 25.4]$.

Problem 9.4.9 — Use of CLT

Let X_i be Uniform(-1,1). Let $Y_i = 20 + 15X_i^2$. Let $W = \frac{1}{100}\sum_{i=1}^{100} Y_i$.

Estimate $\mathrm{P}[W \leq 25.4]$.

$$\begin{aligned}
\mathrm{E}[X_i] &= 0\,, \qquad \mathrm{E}[X_i^2] = \frac{1}{3}\,, \qquad \mathrm{E}[X_i^4] = \int_{-1}^{1} \frac{1}{2}x^4 dx = \frac{1}{5} \\
\mathrm{E}[Y_i] &= 20 + 15\,\mathrm{E}[X_i^2] = 25 \\
\mathrm{E}[Y_i^2] &= 400 + 600\,\mathrm{E}[X_i^2] + 225\,\mathrm{E}[X_i^4] = 645 \\
\mathrm{var}[Y_i] &= \mathrm{E}[Y_i^2] - (\mathrm{E}[Y_i])^2 = 645 - 625 = 20
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}[W] &= \mathrm{E}[Y_i] = 25 \\
\mathrm{var}[W] &= \frac{1}{100}\mathrm{var}[Y_i] = 0.2
\end{aligned}$$

$$\begin{aligned}
\mathrm{P}[W \leq 25.4] = \mathrm{E}\left[\frac{W-25}{\sqrt{0.2}} \leq \frac{25.4-25}{\sqrt{0.2}}\right] = \mathrm{E}[Z \leq 0.8944] &\approx \Phi(0.8944) \\
&= 0.8145
\end{aligned}$$

(Ch.10) The sample mean

The **expected value** is given by

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x) dx.$$

What if $f_X(x)$ is unknown?

- In practice, we estimate $\mathrm{E}[X]$ by averaging independent observations (data samples). But, this sample average is a RV!

(Ch.10) The sample mean

The **expected value** is given by

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x) dx.$$

What if $f_X(x)$ is unknown?

- In practice, we estimate $\mathrm{E}[X]$ by averaging independent observations (data samples). But, this sample average is a RV!

Let $X_1, \cdots, X_n$ be n iid RVs with PDF $f_X(x)$ obtained from n repeated independent trials of an experiment. The **sample mean** of X is then given by the RV

$$M_n(X) = \frac{X_1 + \cdots + X_n}{n}.$$

Expected value and sample mean

Note:

- $\mathrm{E}[X]$ is a number (deterministic)
- $M_n(X) = \frac{X_1 + \cdots + X_n}{n}$ is a function of the RVs $X_1, \cdots, X_n$.
  Hence, $M_n(X)$ is also a RV.

This means we can talk about the expected value $\mathrm{E}[M_n(X)]$ and variance $\mathrm{var}[M_n(X)]$.

Main question to answer: How well does $M_n(X)$ converge to $\mathrm{E}[X]$ as a function of n?

Expected value and sample mean

Because $X_1, \cdots, X_N$ are iid:

$$\begin{aligned} \mathrm{E}\left[M_n(X)\right] &= \mathrm{E}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{1}{n}\left(\mathrm{E}\left[X_1\right] + \cdots + \mathrm{E}\left[X_n\right]\right) \\ &= \frac{1}{n}\left(\mathrm{E}\left[X\right] + \cdots + \mathrm{E}\left[X\right]\right) = \mathrm{E}[X] \end{aligned}$$

$$\begin{aligned} \mathrm{var}\left[M_n(X)\right] &= \frac{1}{n^2}\mathrm{var}\left[X_1 + \cdots + X_n\right] = \frac{\mathrm{var}\left[X_1\right] + \cdots + \mathrm{var}\left[X_n\right]}{n^2} \\ &= \frac{n\,\mathrm{var}[X]}{n^2} = \frac{\mathrm{var}[X]}{n}. \end{aligned}$$

We conclude: as $n \to \infty$, $M_n(X)$ is arbitrarily close to $\mathrm{E}[X]$.

- $M_n(X)$ converges to $\mathrm{E}[X]$. *What does this mean, exactly?*

Problem 10.1.1

$X_1, \cdots, X_n$ is an iid sequence of exponential random variables, each with expected value 5.

(a) What is $\text{var}[M_9(X)]$, the variance of the sample mean based on 9 trials?

(b) What is $P[X_1 > 7]$, the probability that one outcome exceeds 7?

(c) Use the central limit theorem to estimate $P[M_9(X) > 7]$, the probability that the sample mean exceeds 7.

Problem 10.1.1

$X_1, \cdots, X_n$ is an iid sequence of exponential random variables, each with expected value 5.

(a) What is $\text{var}[M_9(X)]$, the variance of the sample mean based on 9 trials?

(b) What is $\text{P}[X_1 > 7]$, the probability that one outcome exceeds 7?

(c) Use the central limit theorem to estimate $\text{P}[M_9(X) > 7]$, the probability that the sample mean exceeds 7.

The X_i have $\mu_X = 5$, $\sigma_X = 5$, $F_X(x) = 1 - e^{-x/5}$.

(a) $\text{var}[M_9(X)] = \dfrac{\sigma_X^2}{9} = \dfrac{25}{9}$

(b) $\text{P}[X_1 > 7] = 1 - \text{P}[X_1 \leq 7] = 1 - F_X(7) = e^{-7/5} \approx 0.247$

(c) $\text{P}[M_9(X) > 7] = 1 - \text{P}[M_9 \leq 7] = 1 - \text{P}[\frac{M_9 - 5}{\text{std}} \leq \frac{7-5}{\text{std}}] \approx 1 - \Phi(\frac{2}{5/3}) \approx 0.1151$

Deviation of a RV from its expected value

How well does $M_n(X)$ converge to $\mathrm{E}[X]$? Consider first:

What is the deviation of a RV X from its expected value: $|X - \mathrm{E}[X]|$?

- **Markov inequality**: If X is nonnegative ($\mathrm{P}[X < 0] = 0$)

$$\mathrm{P}[X \geq c^2] \leq \frac{\mathrm{E}[X]}{c^2}$$

(often inaccurate)

- **Chebyshev inequality**: For a RV X

$$\mathrm{P}[\,|X - \mathrm{E}[X]| \geq c\,] \leq \frac{\mathrm{var}[X]}{c^2}$$

(most often used)

- **Chernoff Bound**:

$$\mathrm{P}[X \geq c] \leq \min_{s \geq 0} e^{-sc} \phi_X(s)$$

(need to know the PDF)

Deviation of a RV from its expected value

How well does $M_n(X)$ converge to $\mathrm{E}[X]$? Consider first:

What is the deviation of a RV X from its expected value: $|X - \mathrm{E}[X]|$?

- **Markov inequality**: If X is nonnegative ($\mathrm{P}[X < 0] = 0$)

$$\mathrm{P}[X \geq c^2] \leq \frac{\mathrm{E}[X]}{c^2}$$ (often inaccurate)

- **Chebyshev inequality**: For a RV X

$$\mathrm{P}[\,|X - \mathrm{E}[X]| \geq c\,] \leq \frac{\mathrm{var}[X]}{c^2}$$ (most often used)

- **Chernoff Bound**:

$$\mathrm{P}[X \geq c] \leq \min_{s \geq 0} e^{-sc} \phi_X(s)$$ (need to know the PDF)

(P10.1.1) Chebyshev:
$\mathrm{P}[M_9(X) > 7] = \mathrm{P}[M_9(X) - 5 > 2] \leq \mathrm{var}[M_9]/4 \approx 0.6944$

Derivations

Markov inequality:

For constant c and a non-negative RV X (i.e., $\mathrm{P}[X < 0] = 0$)

$$\begin{aligned}
\mathrm{E}[X] &= \int_0^\infty x\, f_X(x)\mathrm{d}x = \int_0^{c^2} x\, f_X(x)\mathrm{d}x + \int_{c^2}^\infty x\, f_X(x)\mathrm{d}x \\
&\geq \int_{c^2}^\infty x\, f_X(x)\mathrm{d}x \\
&\geq c^2 \int_{c^2}^\infty f_X(x)\mathrm{d}x \quad \text{since } x \geq c^2
\end{aligned}$$

$$\Rightarrow \quad \mathrm{P}[X \geq c^2] \leq \frac{\mathrm{E}[X]}{c^2}$$

Derivations (cont'd)

Chebyshev inequality:

Using the Markov inequality

$$P[X \geq c^2] \leq \frac{E[X]}{c^2}.$$

- Let $X = |Y - E[Y]|^2$. The Markov inequality then says:

$$P[X \geq c^2] \;=\; P[|Y - E[Y]|^2 \geq c^2] \;\leq\; \frac{E[|Y - E[Y]|^2]}{c^2} \;=\; \frac{\mathrm{var}[Y]}{c^2}.$$

- As $P[|Y - E[Y]|^2 \geq c^2] = P[|Y - E[Y]| \geq c]$, we obtain

$$P[|Y - E[Y]| \geq c] \;\leq\; \frac{\mathrm{var}[Y]}{c^2}$$

  which is the Chebyshev inequality

Derivations (cont'd)

Chernoff bound:

$$\mathrm{P}[X \geq c] = \int_c^\infty f_X(x)\mathrm{d}x = \int_{-\infty}^\infty u(x-c)\, f_X(x)\mathrm{d}x$$

where $u(x)$ is the unit step function.

- Since $u(x-c) \leq e^{s(x-c)}$ for all $s \geq 0$, then

  $$\mathrm{P}[X \geq c] \leq \int_{-\infty}^\infty e^{s(x-c)} f_X(x)\mathrm{d}x = e^{-sc} \int_{-\infty}^\infty e^{sx} f_X(x)\mathrm{d}x = e^{-sc}\phi_X(s)$$

  with $\phi_X(s)$ the moment generating function of X, and any $s \geq 0$.
- To obtain the bound, we can select the s that minimizes $e^{-sc}\phi_X(s)$.

The Chernoff bound is then given by

$$\mathrm{P}[X \geq c] \;\leq\; \min_{s \geq 0} e^{-sc}\phi_X(s).$$

Problem 10.2.6

Use the Chernoff bound to show that the Gaussian(0,1) random variable Z satisfies $P[Z \geq c] \leq e^{-c^2/2}$.

Problem 10.2.6

Use the Chernoff bound to show that the Gaussian(0,1) random variable Z satisfies $\mathrm{P}[Z \geq c] \leq e^{-c^2/2}$.

The $N[0,1]$ random variable Z has MGF $\phi_Z(s) = e^{s^2/2}$. Hence the Chernoff bound for Z is

$$\mathrm{P}[Z \geq c] \leq \min_{s \geq 0} e^{-sc} e^{s^2/2} = \min_{s \geq 0} e^{s^2/2 - sc}$$

We can minimize $e^{s^2/2-sc}$ by minimizing the exponent $s^2/2 - sc$. By setting

$$\frac{\mathrm{d}}{\mathrm{d}s}(s^2/2 - sc) = s - c = 0$$

we obtain $s = c$. At $s = c$, the upper bound is $\mathrm{P}[Z \geq c] \leq e^{-c^2/2}$.

Problem 10.2.6

Use the Chernoff bound to show that the Gaussian(0,1) random variable Z satisfies $\mathrm{P}[Z \geq c] \leq e^{-c^2/2}$.

The $N[0,1]$ random variable Z has MGF $\phi_Z(s) = e^{s^2/2}$. Hence the Chernoff bound for Z is

$$\mathrm{P}[Z \geq c] \leq \min_{s \geq 0} e^{-sc} e^{s^2/2} = \min_{s \geq 0} e^{s^2/2 - sc}$$

We can minimize $e^{s^2/2-sc}$ by minimizing the exponent $s^2/2 - sc$. By setting

$$\frac{\mathrm{d}}{\mathrm{d}s}(s^2/2 - sc) = s - c = 0$$

we obtain $s = c$. At $s = c$, the upper bound is $\mathrm{P}[Z \geq c] \leq e^{-c^2/2}$.

	$c=1$	$c=2$	$c=3$	$c=4$	$c=5$
Chernoff bound	0.606	0.135	0.011	3.35×10^{-4}	3.73×10^{-6}
$Q(c)$	0.159	0.023	0.0013	3.17×10^{-5}	2.87×10^{-7}

Going back to the sample mean...

How well does the sample mean $M_n(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$ converge to $\mathrm{E}[X]$?

Chebyshev inequality applied to $M_n(X)$:

$$\begin{aligned} \mathrm{P}[\,|M_n(X) - \mathrm{E}[X]| \geq c\,] &= \mathrm{P}[\,|M_n(X) - \mathrm{E}[M_n(X)]| \geq c\,] \\ &\leq \frac{\mathrm{var}[M_n(X)]}{c^2} = \frac{\mathrm{var}[X]}{n\,c^2} \end{aligned}$$

This is also known as the (weak) law of large numbers:

- The probability that the sample mean $M_n(X)$ is more than c units away from $\mathrm{E}[X]$ can be made arbitrarily small by making n large enough.

This is called *convergence in probability* (almost sure, a.s.)

Problem 10.3.2

Event A has probability $\mathrm{P}[A] = 0.8$. Let $\hat{P}_n(A)$ denote the relative frequency of event A in n independent trials.
Let X_A denote the indicator random variable for event A.

(a) Find $\mathrm{E}[X_A]$ and $\mathrm{var}[X_A]$.

(b) What is $\mathrm{var}[\hat{P}_n(A)]$.

(c) Use the Chebyshev inequality to find the confidence coefficient $1-\alpha$ such that $\hat{P}_{100}(A)$ is within 0.1 of $\mathrm{P}[A]$.
I.e., find α such that $\mathrm{P}[|\hat{P}_{100}(A) - \mathrm{P}[A]| \leq 0.1] \geq 1-\alpha$.

Problem 10.3.2

Event A has probability $\mathrm{P}[A] = 0.8$. Let $\hat{P}_n(A)$ denote the relative frequency of event A in n independent trials.
Let X_A denote the indicator random variable for event A.

(a) Find $\mathrm{E}[X_A]$ and $\mathrm{var}[X_A]$.
(b) What is $\mathrm{var}[\hat{P}_n(A)]$.
(c) Use the Chebyshev inequality to find the confidence coefficient $1 - \alpha$ such that $\hat{P}_{100}(A)$ is within 0.1 of $\mathrm{P}[A]$.
I.e., find α such that $\mathrm{P}[|\hat{P}_{100}(A) - \mathrm{P}[A]| \leq 0.1] \geq 1 - \alpha$.

(a) Since X_A is a Bernoulli($p = \mathrm{P}[A]$) random variable,

$$\mathrm{E}[X_A] = \mathrm{P}[A] = 0.8\,, \qquad \mathrm{var}[X_A] = \mathrm{P}[A]\,(1 - \mathrm{P}[A]) = 0.16$$

(b) $\hat{P}_n(A) = M_n(X_A) = \frac{1}{n}\sum_{i=1}^{n} X_{A,i}$
$\mathrm{var}[\hat{P}_n(A)] = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{var}[X_{A,i}] = \frac{\mathrm{P}[A](1-\mathrm{P}[A])}{n}$

Problem 10.3.2 (cont'd)

(c) Since $\hat{P}_{100}(A) = M_{100}(X_A)$, we can use the Chebyshev inequality to write

$$\begin{aligned} \mathrm{P}[|\hat{P}_{100}(A) - \mathrm{P}[A]| < c] &\geq 1 - \frac{\mathrm{var}[X_A]}{100c^2} \\ &= 1 - \frac{0.16}{100c^2} = 1 - \alpha \end{aligned}$$

For $c = 0.1$, $\alpha = 0.16/[100(0.1)^2] = 0.16$. Thus, with 100 samples, our confidence coefficient is $1 - \alpha = 0.84$.

Quality of an estimator

The sample mean $M_n(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$ is one example of estimating a model parameter (here, $r = \mathrm{E}[X]$) describing a statistical model.

Also other parameters of a probability model, e.g., the higher order moments $\mathrm{E}[X^2], \mathrm{E}[X^3], \cdots, \mathrm{E}[X^n]$, can be estimated by sample averages.

How to express whether an estimator $\hat{R}$ of a model parameter r is good?

- Bias
- Consistency
- Accuracy (e.g., mean square error)

Unbiased estimator

An estimate $\hat{R}$ of a parameter r is **unbiased** if $\mathrm{E}[\hat{R}] = r$.

Let $\hat{R}_n$ be an estimator of r using observations $X_1, X_2, \cdots, X_n$.

The sequence of estimators $\hat{R}_n$ of a parameter r is **asymptotically unbiased** if

$$\lim_{n \to \infty} \mathrm{E}[\hat{R}_n] = r$$

Consistent estimator

The sequence of estimates $\hat{R}_1, \hat{R}_2, \cdots$ of parameter r is **consistent** if for any $\epsilon > 0$

$$\lim_{n \to \infty} \mathrm{P}\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] = 0$$

I.e., the sequence of estimates $\hat{R}_1, \hat{R}_2, \cdots$ converges in probability.

- Necessary: (asymptotically) unbiased. What else is needed?

Mean square error

The **mean square error** of an estimator $\hat{R}$ of a parameter r is

$$e = \mathrm{E}[(\hat{R} - r)^2]$$

When $\hat{R}$ is unbiased, $\mathrm{E}[\hat{R}] = r$, then

$$e = \mathrm{E}[(\hat{R} - r)^2] = \mathrm{E}[(\hat{R} - \mathrm{E}[\hat{R}])^2] = \mathrm{var}[\hat{R}]$$

Relation MSE, bias and variance

Let $b = \mathrm{E}[\hat{R}] - r$ and $V = \hat{R} - \mathrm{E}[\hat{R}]$, so that $\mathrm{E}[V] = 0$.

$$\begin{aligned} e &= \mathrm{E}[(\hat{R} - r)^2] = \mathrm{E}[(\hat{R} - \mathrm{E}[\hat{R}] + \mathrm{E}[\hat{R}] - r)^2] \\ &= \mathrm{E}[(V + b)^2] = \mathrm{E}[V^2] + 2\mathrm{E}[V]b + b^2 \\ &= \underbrace{\mathrm{E}[V^2]}_{\text{variance}} + \underbrace{b^2}_{\text{bias-squared}} \end{aligned}$$

Mean square error – Theorem 10.8

Theorem: If a sequence of unbiased estimators $\hat{R}_1, \hat{R}_2, \cdots$ of parameter r has a MSE $e_n = \text{var}[\hat{R}_n]$ with $\lim_{n\to\infty} e_n = 0$, then the sequence is consistent.

Proof:

This follows directly from the Chebyshev inequality:

$$P\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] \leq \frac{\text{var}[\hat{R}_n]}{\epsilon^2}$$

Applying Chebyshev for $n \to \infty$:

$$\lim_{n\to\infty} P\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] \leq \lim_{n\to\infty} \frac{\text{var}[\hat{R}_n]}{\epsilon^2} = 0$$

Example

Let N_k be the number of packets per interval of k seconds passing through a router. Assume N_k is Poisson distributed with $\mathrm{E}[N_k] = kr$.

Let $\hat{R}_k = N_k/k$ denote an estimator of the parameter r (number of packets/sec).

(a) Is $\hat{R}_k$ unbiased?

(b) What is the mean square error of $\hat{R}_k$?

(c) Is the sequence $\hat{R}_1, \hat{R}_2, \cdots$ consistent?

Example

Let N_k be the number of packets per interval of k seconds passing through a router. Assume N_k is Poisson distributed with $\mathrm{E}[N_k] = kr$.

Let $\hat{R}_k = N_k/k$ denote an estimator of the parameter r (number of packets/sec).

(a) Is $\hat{R}_k$ unbiased?
(b) What is the mean square error of $\hat{R}_k$?
(c) Is the sequence $\hat{R}_1, \hat{R}_2, \cdots$ consistent?

(a) $\mathrm{E}[\hat{R}_k] = \mathrm{E}[N_k/k] = \mathrm{E}[N_k]/k = r$. Yes, unbiased.
(b) Poisson distributed, so $\mathrm{var}[N_k] = kr$,

$$\mathrm{var}[\hat{R}_k] = \mathrm{var}[N_k/k] = \mathrm{var}[N_k]/k^2 = r/k$$

Unbiased, so the MSE is $e_k = \mathrm{E}[(\hat{R}_k - r)^2] = \mathrm{var}[\hat{R}_k] = r/k$.
(c) $\lim_{k\to\infty} e_k = \lim_{k\to\infty} r/k = 0$, thus consistent according to Theorem 10.8.

Mean square error – Theorem 10.9, 10.11

- The sample mean $M_n(X)$ is an unbiased and consistent estimator of $\mathrm{E}[X]$ (if X has finite variance)
- The sample variance $V_n(X)$ is biased (but asymptotically unbiased).

$$V_n(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - M_n(X))^2$$

The bias happens because $M_n(X)$ also depends on X_i. But

$$V'_n(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M_n(X))^2$$

is an unbiased estimate of $\mathrm{var}[X]$.

Problem 10.4.1

An experimental trial produces random variables X_1 and X_2 with correlation $r = \mathrm{E}[X_1 X_2]$. To estimate r, we perform n independent trials and form the estimate

$$\hat{R}_n = \frac{1}{n} \sum_{i=1}^{n} X_1(i) X_2(i)\,,$$

where $X_1(i)$ and $X_2(i)$ are samples of X_1 and X_2 on trial i.
Show that if $\mathrm{var}[X_1 X_2]$ is finite, then $\hat{R}_1, \hat{R}_2, \cdots$ is an unbiased, consistent sequence of estimates of r.

Problem 10.4.1

An experimental trial produces random variables X_1 and X_2 with correlation $r = \mathrm{E}[X_1 X_2]$. To estimate r, we perform n independent trials and form the estimate

$$\hat{R}_n = \frac{1}{n} \sum_{i=1}^{n} X_1(i) X_2(i),$$

where $X_1(i)$ and $X_2(i)$ are samples of X_1 and X_2 on trial i.
Show that if $\mathrm{var}[X_1 X_2]$ is finite, then $\hat{R}_1, \hat{R}_2, \cdots$ is an unbiased, consistent sequence of estimates of r.

Let $Y = X_1 X_2$, and for the ith trial, let $Y_i = X_1(i) X_2(i)$.
Then $\hat{R}_n = M_n(Y)$, the sample mean of random variable Y. By Theorem 10.9, $M_n(Y)$ is unbiased.

Since $\mathrm{var}[Y] = \mathrm{var}[X_1 X_2] < \infty$, Theorem 10.11 tells us that $M_n(Y)$ is a consistent sequence.

To do for this lecture:

- Read chapter 6.2, 6.5, 9 and 10
- Make some of the indicated exercises:
  6.2.1, 6.2.5, 6.2.7, 9.2.1, 9.2.3, 9.3.3, 9.3.5, 9.3.7,
  10.2.1, 10.2.3, 10.2.5, 10.3.1