Moments

- Let $X$ be a continuous random variable with density function $f(x)$
- $k^{th}$ moment of $X$, if it exists, is:
  \[ E[X^k] = \int_{-\infty}^{\infty} x^k f(x) \, dx \]
- $\mu = E[X]$ is the first moment of $X$
- $k^{th}$ central moment of $X$ is:
  \[ \mu_k = E[(X - \mu)^k] = \int_{-\infty}^{\infty} (x - \mu)^k f(x) \, dx \]
- $\sigma = \sqrt{E[(X - \mu)^2]}$ standard deviation is the square root of the second central moment
- $k^{th}$ standardized moment of $X$ is:
  \[ \tilde{\mu}_k = \frac{\mu_k}{\sigma^k} = E \left[ \left( \frac{X - \mu}{\sigma} \right)^k \right] \]
Skewness

- $\tilde{\mu}_1 = E[(X-\mu)]/\sigma = 0$ since $E[X - \mu] = 0$
- $\tilde{\mu}_2 = E[(X-\mu)^2]/\sigma^2 = 1$ since $\sigma^2 = E[(X - \mu)^2]$
- $\tilde{\mu}_3 = E[(X-\mu)^3]/\sigma^3$ [(Pearson’s moment) coefficient of skewness]
- Skewness indicates direction and magnitude of a distribution’s deviation from symmetry

E.g., for $X \sim \text{Exp}(\lambda)$, $\tilde{\mu}_3 = 2$
Kurtosis

- \( \tilde{\mu}_4 = E[\left(\frac{X-\mu}{\sigma}\right)^4] \)  
- For \( X \sim N(\mu, \sigma) \), \( \tilde{\mu}_4 = 3 \)
- Kurtosis is a measure of the dispersion of \( X \) around the two values \( \mu \pm \sigma \)

\( \tilde{\mu}_4 \) > 3 Leptokurtic (slender) distribution has fatter tails. May have outlier problems.
\( \tilde{\mu}_4 \) < 3 Platykurtic (broad) distribution has thinner tails

See R script
Functions of two or more random variables: expectation

- $V = \pi HR^2$ be the volume of a vase of height $H$ and radius $R$
- $g(H, R) = \pi HR^2$ is a random variable (function of random variables)
- $P_V(V = 3) = P_{HR}(\pi HR^2 = 3)$
- How to calculate $E[V]$?

Two-dimensional change-of-variable formula. Let $X$ and $Y$ be random variables, and let $g: \mathbb{R}^2 \to \mathbb{R}$ be a function. If $X$ and $Y$ are discrete random variables with values $a_1, a_2, \ldots$ and $b_1, b_2, \ldots$, respectively, then

$$E[g(X, Y)] = \sum_i \sum_j g(a_i, b_j)P(X = a_i, Y = b_j).$$

If $X$ and $Y$ are continuous random variables with joint probability density function $f$, then

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y) \, dx \, dy.$$

If $H \perp \perp R$:

$$E[V] = E[\pi HR^2] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \pi hr^2 f_H(h)f_R(r) \, dh \, dr$$
**Theorem.** For $X$ and $Y$ random variables, and $s, t \in \mathbb{R}$:

$$E[rX + sY + t] = rE[X] + sE[Y] + t$$

**Proof.** (discrete case)

$$E[rX + Ys + t] = \sum_a \sum_b (ra + sb + t)P(X = a, Y = b)$$

$$= \left( r \sum_a \sum_b aP(X = a, Y = b) \right) + \left( s \sum_a \sum_b bP(X = a, Y = b) \right) + \left( t \sum_a \sum_b P(X = a, Y = b) \right)$$

$$= \left( r \sum_a aP(X = a) \right) + \left( s \sum_b bP(Y = b) \right) + t = rE[X] + sE[Y] + t$$

**Corollary.** $E[a_0 + \sum_{i=1}^n a_iX_i] = a_0 + \sum_{i=1}^n a_iE[X_i]$  

**Corollary.** $X \leq Y$ implies $E[X] \leq E[Y]$  

**Proof.** $Z = Y - X \geq 0$ implies $E[Z] = E[Y] - E[X] \geq 0$, i.e., $E[Y] \geq E[X]$. 
Applications

- **Expectation of some discrete distributions**
  - $X \sim Ber(p) \quad E[X] = p$
  - $X \sim Bin(n, p) \quad E[X] = n \cdot p$
    - Because $X = \sum_{i=1}^{n} X_i$ for $X_1, \ldots, X_n \sim Ber(p)$
  - $X \sim Geo(p) \quad E[X] = \frac{1}{p}$
  - $X \sim NBin(n, p) \quad E[X] = \frac{n \cdot (1 - p)}{p}$
    - Because $X = \sum_{i=1}^{n} X_i - n$ for $X_1, \ldots, X_n \sim Geo(p)$

- **Expectation of some continuous distributions**
  - $X \sim Exp(\lambda) \quad E[X] = \frac{1}{\lambda}$
  - $X \sim Erl(n, \lambda) \quad E[X] = \frac{n}{\lambda}$
    - Because $X = \sum_{i=1}^{n} X_i$ for $X_1, \ldots, X_n \sim Exp(\lambda)$
**Theorem.** For $X \perp \perp Y$, we have: $E[XY] = E[X]E[Y]$  

**Proof.** $X \perp \perp Y$ implies $X \perp \perp 1/Y$. By theorem above:


because by Jensen’s inequality $E[1/Y] \geq 1/E[Y]$ since $1/y$ is convex for $y \geq 0$. □

**Corollary.** For $X \perp \perp Y$ and $Y \geq 0$, we have: $E[X/Y] \geq E[X]/E[Y]$  

**Proof.** $X \perp \perp Y$ and $Y \geq 0$ implies $X \perp \perp 1/Y$. By theorem above:


because by Jensen’s inequality $E[1/Y] \geq 1/E[Y]$ since $1/y$ is convex for $y \geq 0$. □

**Exercise at home.** Show that $E[X/Y] = E[X]/E[Y]$ is a false claim.
**Law of iterated/total expectation**

### Conditional expectation

\[ E[X|Y = b] = \sum_i a_i p(a_i|b) \]

\[ E[X|Y = y] = \int_{-\infty}^{\infty} xf(x|y)dx \]

**Theorem.** (Law of iterated/total expectation)

\[ E_Y[E[X|Y]] = E[X] \]

**Proof.** (for \(X, Y\) discrete random variables)

\[ E_Y[E[X|Y]] = \sum_j \sum_i a_i p_{X|Y}(a_i|b_j) p_Y(b_j) = \sum_j \sum_i a_i p_{XY}(a_i, b_j) = \sum_i a_i p_X(a_i) = E[X] \]

**Example** (cfr the example from Lesson 1 on the Law of total probability)

- Factory 1’s light bulbs working hours \(\sim \text{Exp}(1/1000)\)
- Factory 2’s light bulbs working hours \(\sim \text{Exp}(1/2000)\)
- Factory 1 supplies 60% of the total bulbs on the market and Factory 2 supplies 40% of it.
- **What is the average work hour of a light bulb on the market?**
Variance of the sum and covariance

\[ \text{Var}(X + Y) = E[(X + Y - E[X + Y])^2] = E[((X - E[X]) + (Y - E[Y]))^2] \]
\[ = E[(X - E[X])^2] + E[(Y - E[Y])^2] + 2E[(X - E[X])(Y - E[Y])] \]
\[ = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \]

Covariance

The covariance \( \text{Cov}(X, Y) \) of two random variables \( X \) and \( Y \) is the number:

\[ \text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \]
Covariance

**Theorem.** \( \text{Cov}(X, Y) = E[XY] - E[X]E[Y] \)

- If \( X \) and \( Y \) are independent \( (X \perp \!\!\!\!\perp Y) \):
  \[
  \text{Cov}(X, Y) = 0 \quad \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)
  \]

- But there are \( X \) and \( Y \) uncorrelated (ie., \( \text{Cov}(X, Y) = 0 \)) that are dependent!

- Variances of some discrete distributions
  - \( X \sim \text{Ber}(p) \quad \text{Var}(X) = p(1 - p) \)
  - \( X \sim \text{Bin}(n, p) \quad \text{Var}(X) = np(1 - p) \)
    - Because \( X = \sum_{i=1}^{n} X_i \) for \( X_1, \ldots, X_n \sim \text{Ber}(p) \) and independent
  - \( X \sim \text{Geo}(p) \quad \text{Var}(X) = \frac{1-p}{p^2} \)
  - \( X \sim \text{NBin}(n, p) \quad \text{Var}(X) = n\frac{1-p}{p^2} \)
    - Because \( X = \sum_{i=1}^{n} X_i - n \) for \( X_1, \ldots, X_n \sim \text{Geo}(p) \) and independent

- Variances of some continuous distributions
  - \( X \sim \text{Exp}(\lambda) \quad \text{Var}(X) = \frac{1}{\lambda^2} \)
  - \( X \sim \text{Erl}(n, \lambda) \quad \text{Var}(X) = \frac{n}{\lambda^2} \)
    - Because \( X = \sum_{i=1}^{n} X_i \) for \( X_1, \ldots, X_n \sim \text{Exp}(\lambda) \) and independent
Covariance and covariance matrix

Let $X$ and $Y$ be two random variables. Then

$$
\text{Cov}(rX + sY + t) = rt \text{Cov}(X, Y)
$$

for all numbers $r, s, t,$ and $u.$

• Hence, $\text{Var}(rX + sY + t) = r^2 \text{Var}(X) + s^2 \text{Var}(Y) + 2rs\text{Cov}(X, Y)$

• Covariance depends on the unit of measure!

• Bivariate Normal/Gaussian distribution:

$$(X, Y) \sim N((\mu_x, \mu_y), \begin{pmatrix}
\sigma_x^2 & \sigma_{xy} \\
\sigma_{xy} & \sigma_y^2
\end{pmatrix})$$

where marginals are $X \sim N(\mu_x, \sigma_x^2),$ $Y \sim N(\mu_y, \sigma_y^2),$ and $\text{Cov}(X, Y) = \sigma_{xy}$

• Covariance matrix $\Sigma_{ij} = \text{Cov}(X_i, X_j)$ for a vector $X = (X_1, \ldots, X_n)$ of r.v.'s

See R script lesson 08
Correlation coefficient

**Definition.** Let $X$ and $Y$ be two random variables. The correlation coefficient $\rho(X, Y)$ is defined to be 0 if $\text{Var}(X) = 0$ or $\text{Var}(Y) = 0$, and otherwise

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}.$$ 

- Correlation coefficient is *dimensionless* (not affected by change of units)
  - E.g., if $X$ and $Y$ are in Km, then $\text{Cov}(X, Y)$, $\text{Var}(X)$ and $\text{Var}(Y)$ are in Km$^2$
- Moreover: $-1 \leq \rho(X, Y) \leq 1$
  - The bounds are derived from the **Cauchy–Schwarz's inequality**:

$$E[|XY|] \leq \sqrt{E[X^2]} \sqrt{E[Y^2]}$$

**Proof.** For any $u, w \in \mathbb{R}$, we have $2|uw| \leq u^2 + w^2$. Therefore, $2|UV| \leq U^2 + W^2$ for r.v.'s $U$ and $V$. By defining $U = X/\sqrt{E[X^2]}$ and $W = Y/\sqrt{E[Y^2]}$ ($\ast$), we have

$$2 \cdot \frac{|XY|}{\sqrt{E[X^2]} \sqrt{E[Y^2]}} \leq \frac{X^2}{E[X^2]} + \frac{Y^2}{E[Y^2]}.$$ 

Taking the expectations, we conclude:

$$2 \cdot E[|XY|]/\sqrt{E[X^2]} \sqrt{E[Y^2]} \leq 2. \quad \ast$$

($\ast$) The case $E[X^2] = 0$ or $E[Y^2] = 0$ is left as an exercise. □
Bivariate Normal/Gaussian distribution

\[(X, Y) \sim N((\mu_x, \mu_x), \left( \begin{array}{cc} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{array} \right))\]

where marginals are \(X \sim N(\mu_x, \sigma_x^2), Y \sim N(\mu_y, \sigma_y^2)\), and \(\text{Cov}(X, Y) = \sigma_{xy}\)

- Since \(\sigma_{xy} = \rho(X, Y) \cdot \sigma_x \cdot \sigma_y\):

\[(X, Y) \sim N((\mu_x, \mu_x), \left( \begin{array}{cc} \sigma_x^2 & \rho(X, Y) \cdot \sigma_x \cdot \sigma_y \\ \rho(X, Y) \cdot \sigma_x \cdot \sigma_y & \sigma_y^2 \end{array} \right))\]

- Density of \(N((0, 0), (1, \sigma_{xy}, \sigma_{xy}, 1))\):

\[f(x, y) = \frac{1}{2\pi \sqrt{1 - \sigma_{xy}^2}} e^{-\frac{1}{2(1-\sigma_{xy}^2)}(x^2+y^2-2\sigma_{xy}xy)}\]

- Useful facts for \((X, Y)\) bivariate Normal:
  - \((X, Y)\) bivariate Normal iff \(aX + bY\) is Normal for any \(a, b \in \mathbb{R}\)
  - for \((X, Y)\) bivariate Normal: \(\rho(X, Y) = 0\) iff \(X \perp \perp Y\), i.e., uncorrelation equals independence
Sum of independent random variables (repetita iuvant)

- See Lesson 04 and Lesson 08 for convolution formulas

**Examples:**

- For $X \sim Bin(n, p)$ and $Y \sim Bin(m, p)$, $Z \sim Bin(n + m, p)$
- For $X \sim Geo(p)$ (days radio 1 breaks) and $Y \sim Geo(p)$ (days radio 2 breaks):

  
  \[ p_Z(X + Y = k) = \sum_{l=1}^{k-1} p_X(l) \cdot p_Y(k - l) = (k - 1)p^2(1 - p)^{k-2} \]
Sum of independent Normal random variables

- See Lesson 04 and Lesson 08 for convolution formulas

**Theorem.** If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ and $X \perp Y$, then:

$$Z = X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

**Proof.** See [T, Sect. 11.2]

- In general: $Z = rX + sY + t \sim N(r\mu_X + s\mu_Y + t, r^2\sigma_X^2 + s^2\sigma_Y^2)$

- The converse of the theorem also holds: [Lévy-Cramér theorem]
  - If $X \perp Y$ and $Z = X + Y$ is normally distributed, then $X$ and $Y$ follow a normal distribution.
Extremes of independent random variables

The distribution of the maximum. Let $X_1, X_2, \ldots, X_n$ be $n$ independent random variables with the same distribution function $F$, and let $Z = \max\{X_1, X_2, \ldots, X_n\}$. Then

$$F_Z(a) = (F(a))^n.$$ 

- $P(Z \leq a) = P(X_1 \leq a, \ldots, X_n \leq a) = \prod_{i=1}^n P(X_i \leq a) = ((F(a))^n$
- Example: maximum water level over 365 days assuming water level on a day is $U(0,1)$
- Example: maximum of two rolls of a die with 4 sides

The distribution of the minimum. Let $X_1, X_2, \ldots, X_n$ be $n$ independent random variables with the same distribution function $F$, and let $V = \min\{X_1, X_2, \ldots, X_n\}$. Then

$$F_V(a) = 1 - (1 - F(a))^n.$$ 

- $P(V \leq a) = 1 - P(X_1 > a, \ldots, X_n > a) = 1 - \prod_{i=1}^n (1 - P(X_i \leq a)) = 1 - ((1 - F(a))^n$
Product and quotient of independent random variables

**Product of Independent Continuous Random Variables.** Let $X$ and $Y$ be two independent continuous random variables with probability densities $f_X$ and $f_Y$. Then the probability density function $f_Z$ of $Z = XY$ is given by

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y \left( \frac{z}{x} \right) f_X(x) \frac{1}{|x|} \, dx$$

for $-\infty < z < \infty$.

**Quotient of Independent Continuous Random Variables.** Let $X$ and $Y$ be two independent continuous random variables with probability densities $f_X$ and $f_Y$. Then the probability density function $f_Z$ of $Z = X/Y$ is given by

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(zx) f_Y(x) |x| \, dx$$

for $-\infty < z < \infty$.

- $X, Y \sim N(0, 1)$ independent, $Z = X/Y \sim \text{Cau}(0, 1)$ where:

$$f_Z(x) = \frac{1}{\pi(1 + x^2)}$$