# ST5201: Basic Statistical Theory
# Chapter 4: Expected Values

Choi, Yunjin
stachoiy@nus.edu.sg

Department of Statistics and Applied Probability
National University of Singapore (NUS)

12th September, 2017

# Announcement

- Midterm on 3rd October (in class):
  - From lecture 1 to lecture 5.
  - One sheet of two-sided A4 allowed
  - A non-programmable calculator is allowed and might be necessary (e.g., Ti-84 is NOT allowed)

- Assignment 2 released:
  - Due on 19th September

# Outline

- Introduction
- Expected Value of a Random Variable
- Variance & Standard Deviation
- Covariance & Correlation
- Conditional Expectation
- Moment-Generating Function

Learning Outcomes

- Questions to Address: What is the expected value $E(X)$ $\star$ How to calculate $E(X)$ of various r.v.'s $\star$ Theoretical properties & results of expectations $\star$ What a variance/covariance is $\star$ What a conditional expectation is $\star$ & What a moment-generating function (mgf) & its utility are $\star$ Determination of distribution/$k$th moment by mgf $\star$ How to obtain mgf of a linear transformation of a r.v./a sum of independent r.v.'s

## Concept & Terminology

- expected value/expectation/mean $\star$ longrun average
- Markov's/Chebyshevs inequality $\star$ expectation of a function of a r.v./a function of $\leq$2 r.v.'s/a linear combination of r.v.'s/$\leq$ 2 independent r.v.'s
- variance & standard deviation $\star$ variability/spread of a r.v. $\star$ variance of a linear transformation of a r.v./a sum of independent r.v.s $\star$ covariance of 2 r.v.'s $\star$ covariance of 2 sums of r.v.'s
- conditional expectation $\star$ law of total expectation
- moment-generating function (mgf) $\star$ $k$th moment $\star$ mgf of a linear transformation of a r.v./a sum of independent r.v.'s

## Mandatory Reading

Textbook: Section 4.1 – Section 4.5

# A Representative of a r.v.

Re-visit of r.v.:

- **Recall**: In prob & stat, an experiment of which a numerical event is concerned is modeled/described by a r.v. $X$, which is characterized by a density (pmf/pdf)

- **Recall**: In general, $> 1$ possible values for any r.v. $X$

  Different values are observed in different occasions

  Which value is observed/realized is governed by the density

- **"1 number" versus "A function"**: More handy to have 1 particular informative value/number to summarize the density (a function)
  - roughly understand the r.v. in some sense
  - compare between different r.v.'s

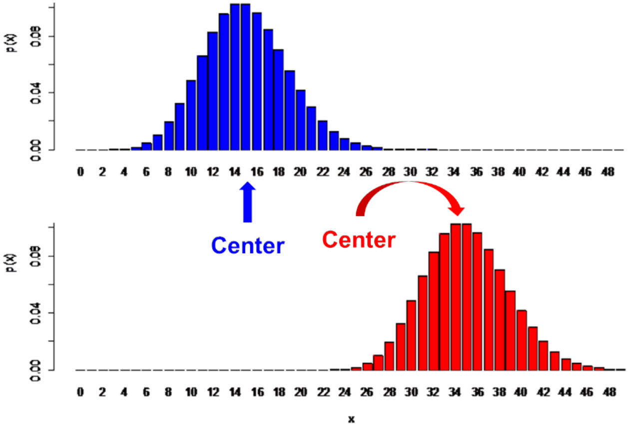
- **Roulette Example**: Suppose a r.v. $X$ represents the (monetary) outcome of a \$1 bet on a single number ("straight up" bet). If the bet wins (which happens with probability $\frac{1}{38}$), the payoff is \$35; otherwise the player loses the bet. The **expected profit** from such a bet will be

$$E(\text{gain from \$1 bet}) = -\$1 \cdot \frac{37}{38} + \$35 \cdot \frac{1}{38} = -\$0.0526.$$

- **Fair Six-sided Die**: Let $X$ represent the outcome of a roll of a fair six-sided die. More specifically, $X$ will be the number of pips showing on the top face of the die after the toss. The **expectation of $X$** is

$$1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5.$$

*1 common way* to obtain such an informative number in regards to the “center” of a distribution is based on the idea of an **average**



**Definition**

For a discrete r.v. $X$ with pmf $p(x)$, the *expected value/expectation/mean of $X$*, denoted by $\mu/\mu_X/E(X)$, is

$$E(X) = \sum_i x_i p(x_i),$$

provided that $\sum_i |x_i| p(x_i) < \infty$. If the sum diverges, the expectation is undefined.

- The range of $X$ is $\{x_1, x_2, \ldots\}$
- Regarded as the center of mass of $p(x)$
- A *weighted average* of all possible values that $X$ can take on, each value being weighted by the prob that $X$ assumes it
- Possible that a discrete r.v. does not have expectation

**Definition**

For a cont. r.v. $X$ with pdf $f(x)$, the *expected value/expectation/mean of $X$*, denoted by $\mu/\mu_X/E(X)$, is

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx,$$

provided that $\int_{-\infty}^{\infty} |x| f(x) dx < \infty$. If the integral diverges, the expectation is undefined.

- The range of $X$ is on an interval
- It is possible that a cont. r.v. does not have an expectation

# Example: Expected Value of a r.v.

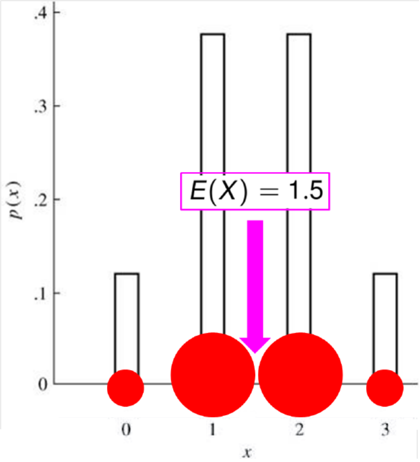
Consider the **Toss 3 fair coins** example:

The pmf of $X$ is given by

$$p(x) = \begin{cases} .125, & x = 0, 3 \\ .375, & x = 1, 2 \\ 0, & \text{otherwise} \end{cases}$$

Hence,

$$\begin{aligned} E(X) &= (0+3)(.125) + (1+2)(.375) \\ &= 1.5 \end{aligned}$$



# Example: Expected Value & Probability

Consider $I_A$, the *indicator r.v. of $A$*, for any event of interest $A \subset \Omega$, defined by

$$I_A = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{if } A^c \text{ occurs} \end{cases}$$

Clearly, $I_A$ is a discrete r.v. taking on 2 values, 1 & 0, with pmf summarized by $p(1) = P(I_A = 1) = P(A)$ &
$p(0) = P(I_A = 0) = P(A^c) = 1 - P(A) \Rightarrow I_A \sim Ber(P(A))$

According to the definition of Expectation,

$$E(I_A) = (1)[P(A)] + (0)[1 - P(A)] = P(A)$$

Theoretically, one can always

**view any prob as an expectation or vice versa**

# Example: Expectation of a Bin(n, p) r.v.

Consider $X \sim Bin(n,p)$ with $p(x) = \binom{n}{x} p^x q^{n-x}$, $x = 0, 1, \cdots, n$

$$
\begin{aligned}
E(X) &= \sum_{x=0}^{n} xp(x) = \sum_{x=1}^{n} (x) \frac{n!}{x!(n-x)!} q^x q^{n-x} \\
&= (np) \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x}, \quad [\text{let } y = x-1] \\
&= (np) [\sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-1-y)!} p^y q^{n-1-y}] \\
&= np
\end{aligned}
$$

since the latter sum is that of all probs of a $Bin(n-1,p)$ r.v.
Remark: Of course, for $Y \sim Ber(p)$, one can work out $E(Y) = p$ either from definition or by setting $n = 1$ in this result.

# Example: Expectation of Unif. & Normal Dist.

- Consider $X \sim U(a,b)$ with $f(x) = \frac{1}{b-a}$ on $[a,b]$ and 0 otherwise.

$$
\begin{aligned}
E(X) &= [\int_{-\infty}^{a} + \int_{a}^{b} + \int_{b}^{\infty}] x f(x)\, dx = 0 + \int_{a}^{b} (x) \frac{1}{b-a}\, dx + 0 \\
&= \frac{1}{b-a} \int_{a}^{b} x dx = \frac{1}{b-a} [\frac{x^2}{2}]|_{a}^{b} = \frac{1}{b-a} [\frac{b^2 - a^2}{2}] = \frac{a+b}{2}
\end{aligned}
$$

It is intuitive that the "center" of the uniform distribution on $[a,b]$ is at the mid-point of the boundaries $a$ & $b$

- Consider $X \sim N(\mu, \sigma^2)$. The first parameter $\mu$ is called the mean parameter of the normal r.v. because $E(X) = \mu$. Refer to the proof in textbook, Page 119.

# Example: r.v. Without Expectation

- Suppose r.v. $X$ takes value 1, -2, 3, -4, ..., with respective prob. $\frac{c}{1^2}$, $\frac{c}{2^2}$, $\frac{c}{3^2}$, $\frac{c}{4^2}$, ..., where $c = 6/\pi^2$ is a normalizing constant that ensures the prob. sum up to 1.
  Then the infinite sum is

$$\sum_{i=1}^{\infty} |x_i| p_i = c \sum_{i=1}^{\infty} \frac{1}{i} = \infty$$

  $\Rightarrow E[X]$ does not exist.

- Cauchy Distribution:

$$f(x) = \frac{1}{\pi} 1 + x^2, \qquad -\infty < x < \infty.$$

  Note that

$$\int_{-\infty}^{\infty} |x| f(x)\, dx = 2 \int_{0}^{\infty} \frac{1}{\pi} \frac{x}{1 + x^2}\, dx = \infty$$

  $\Rightarrow$ its expectation does not exist.

An expected value $E(X)$ can be interpreted as **a long-run average**
When the same experiment is repreated/replicated for a log of times independently

**What is "expected" to be the value of the r.v. of interest?**

- **Toss 3 fair coins**: When we repeatedly toss 3 fair coins for a lot of times, the average of all "# of heads" would be close to 1.5
- **Indicator r.v.**: When we repeatedly perform any experiment for a lot of times independently, the proportion of times of occurrence of $A$ would be close to $E(I_A) = P(A)$; this supports

  empirical determination of any prob.

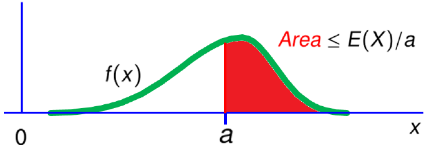
- **The uniform r.v.** $X \sim U(a, b)$: When we repeatedly select a value on $[a, b]$ randomly for a lot of times, the average of all the selected #'s would be close to $E(X) = (a + b)/2$

- There are many other "representatives" such as the *median* & *mode* of a distribution which can serve more or less the same purpose as an indicator of the "center" of a distribution
- $E(X)$ **stands out** among others: There exist *many nice theoretical properties & results about $E(X)$ for any r.v. $X$*

**Markov's Inequality**

For a nonnegative r.v. $X$ (i.e., $P(X \geq 0) = 1$) & $a > 0$,

$$P(X \geq a) \leq \frac{E(X)}{a}.$$



# Proof: Markov's Inequality

Take a cont. r.v. $X$ as an example.

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f(x)\,dx \\
&= \int_{-\infty}^{a} x f(x)\,dx + \int_{a}^{\infty} x f(x)\,dx \\
&\geq \int_{0}^{a} 0 f(x)\,dx + \int_{a}^{\infty} a f(x)\,dx \\
&= 0 + a\int_{a}^{\infty} f(x)\,dx = aP(X \geq a)
\end{aligned}
$$

$\Rightarrow P(X \geq a) \leq E(X)/a$.

- Nonnegative is required.
  For a nonpositive r.v. $X$, similar results can be derived with different direction of inequality.
- Used to control tail prob. & derive more inequalities in prob

# Example: Markov's Inequality

- Consider $X \sim G(\alpha, \lambda)$ with $E(X) = \frac{\alpha}{\lambda}$. We can obtain some bounds of certain gamma probs:

$$\begin{aligned} P(X \geq \alpha) &\leq \frac{1}{\lambda} \\ P(X \geq \frac{2\alpha}{\lambda}) &\leq \frac{1}{2} \end{aligned}$$

- Consider $X \sim B(a, b)$ with $E(X) = \frac{a}{a+b}$. We can obtain some bounds of certain beta probs:

$$\begin{aligned} P(a \leq X < 1) &\leq \frac{1}{a+b} \\ P(\frac{3a}{a+b} \leq X < 1) &\leq \frac{1}{3} \end{aligned}$$

- We discuss how to find the distribution of the r.v. $Y = g(X)$ for a known function $g$ based on knowing the density of $X$, **how about the expectation of $Y$?**

**Expectation of a Function of a r.v.**

Suppose that $Y = g(X)$ for a known function: $g : \mathbb{R} \to \mathbb{R}$

- If $X$ is a discrete r.v. with pmf $p_X(x)$, then $E(Y) = \sum_x g(x)p_X(x)$ provided that $\sum_x |g(x)|p_X(x) < \infty$
- If $X$ ia a cont. r.v. with pdf $f_X(x)$, then $E(Y) = \int_{-\infty}^{\infty} g(x)f_X(x)dx$ provided that $\int_{-\infty}^{\infty} |g(x)|f_X(x)dx < \infty$

- In general, $E(g(X)) \neq g(E(X))$
- Note: $p_X(x)/f_X(x)$ is available $\Rightarrow$ No need to derive $p_Y(y)/f_Y(y)$ in computing $E(Y)$ (with $Y = g(X)$) based on definition of expectation

For a r.v. $X$ with pmf given by $p(x) = \begin{cases} .125, & x = 0, 3 \\ .375, & x = 1, 2 \\ 0, & \text{otherwise} \end{cases}$. Find the pmf of $X^2$, $E(X^2)$ & $E(2X+3)$

**Solution:**

1. The pmf of $Y = X^2$ is defined by
$$P(Y = y) = P(X = \sqrt{y}) = \begin{cases} .125, & y = 0, 9 \\ .375, & y = 1, 4 \\ 0, & \text{otherwise} \end{cases}$$
2. $E(X^2) = \sum_{x=0}^{3} x^2 p(x) = 0^2(.125) + 1^2(.375) + 2^2(.375) + 3^2(.125)$
$= 3 \ (\neq (E(X))^2 = 1.5^2 = 2.25)$
3. $E(2X+3) = \sum_{x=0}^{3} (2x+3)p(x)$
$= [2(0)+3](.125)+[2(1)+3](.375)+[2(2)+3](.375)+[2(3)+3](.125)$
$= 3(.125) + 5(.375) + 7(.375) + 9(.125) = 6$

# Example: Expetation of a Function of a r.v.

Consider finding the expected value of a *chi–square r.v. with 1 degree of freedom*, $Y \sim \chi^2_1$ defined in Example (Ch.2). As $Y = g(Z) = Z^2$ where $Z \sim N(0,1)$ with $f_Z(z)$ as its density,

$$
\begin{aligned}
E(Y) &= \int_{-\infty}^{\infty} g(z) f_Z(z)\, dz \\
&= \int_{-\infty}^{\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz = 2\int_0^{\infty} \frac{z^2}{\sqrt{2\pi}} e^{-z^2/2}\, dz \quad [\text{let } y = \frac{z^2}{2}] \\
&= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} (2y)^{1/2} e^{-y}\, dy = \frac{2}{\sqrt{\pi}} \int_0^{\infty} y^{3/2-1} e^{-y}\, dy \\
&= \frac{2}{\Gamma(1/2)} \Gamma(3/2) = 1
\end{aligned}
$$

Even without knowing that $\chi^2_1 \equiv G(1/2, 1/2)$ or its density, it is still possible to find its expectation through the transformation $Y = Z^2$

# Expectation of a Function of r.v.'s I

- We discuss how to find the joint distribution of the r.v.'s $U = g_1(X, Y)$ & $V = g_2(X, Y)$ for fixed functions $g_1$ & $g_2$ based on knowing the joint density $f_{X,Y}$

**how about the expectation of the r.v. $Z = g(X, Y)$?**

### Expectation of a Function of 2 r.v.'s

- If $X$ & $Y$ have a joint pmf $p(x, y)$, then

$$E(Z) = E[g(X, Y)] = \sum_y \sum_x g(x, y)p(x, y)$$

- If $X$ & $Y$ have a joint pdf $f(x, y)$, then

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

- Note: No need to find $p_Z/f_Z$ with $p_{X,Y}/f_{X,Y}$ known

# Expectation of a Function a r.v.'s II

- One can generalize the previous result for *expectations of a function of $n \geq 2$ r.v.'s*, $E(Z) = E[g(X_1, \cdots, X_n)]$, by replacing $g$ & the joint pmf/pdf accordingly, & replacing the double sum/integral by the corresponding $n$-fold sum/integral

**Expectation of a Function of $n \geq 2$ r.v.'s**

- If $X_1, \cdots, X_n$ have a joint pmf $p(x_1, \cdots, x_n)$, then

$$E[g(X_1, \cdots, X_n)] = \sum_{x_1} \cdots \sum_{x_n} g(x_1, \cdots, x_n) p(x_1, \cdots, x_n)$$

- If $X_1, \cdots, X_n$ have a joint pdf $f(x_1, \cdots, x_n)$, then

$$E[g(X_1, \cdots, X_n)] = \int \cdots \int g(x_1, \cdots, x_n) f(x_1, \cdots, x_n) dx_1 \cdots dx_n$$

### Expectation of Linear Combination of r.v.'s

Suppose that $X_1, \cdots, X_n$ are $n \geq 1$ r.v.'s with expectations $E(X_i)$. For fixed constants $a$, $b_1, \cdots, b_n \in \mathbb{R}$, the *expected value of* $Y = a + \sum_{i=1}^n b_i X_i$ is

$$E(Y) = a + \sum_{i=1}^{n} b_i E(X_i)$$

- The independence between $X_1, \cdots, X_n$ are not assumed. Especially, it holds even if $X_1 = X_2 = \cdots = X_n = X$
- For linear combination $g(\cdot)$, $E(g(X)) = g(E(X))$.
- e.g. on Page 20, $E(2X+3)$ can be alternatively computed as $E(2X+3) = 2E(X) + 3 = 2(1.5) + 3 = 6$

1. Consider $Y \sim Bin(n, p)$. From its construction, $Y$ is *the sum of the # of successes in n indept Bernoulli trials*, *i.e.*,

$$Y = X_1 + X_2 + \cdots + X_n$$

where $X_i \sim Ber(p)$ with $E(X_i) = p$. Hence,
$E(Y) = p + p + \cdots + p = np$

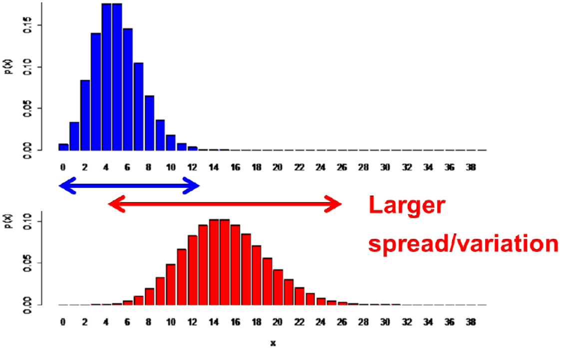
2. Consider $Y \sim NegBin(r, p)$. We can represent it as

$$Y = X_1 + X_2 + \cdots + X_r$$

where $X_1$ is the # of trials required to obtain the 1st success, $X_2$ the # of additional trials until the 2nd success is obtained, $X_3$ the # of additional trials after the 2nd success until the 3rd success is obtained, & so on. That is, $X_i$ represents the # of additional trials required, after the $(i-1)$st success, until a total of $i$ successes is amassed. Clearly, $X_i \sim Geo(p)$ with $E(X_i) = 1/p$. Hence,
$E(Y) = 1/p + 1/p + \cdots + 1/p = r/p$

Another common way to obtain an informative # in describing/summarizing a r.v. concerns about the *variation, or spread*, i.e.,

**how dispersed the distribution is about its center**



# Variance & Standard Deviation I

**Definition**

For any r.v. $X$ with mean $\mu < \infty$, the *variance of $X$* is defined by

$$\text{Var}(X) = E[(X-\mu)^2] \geq 0$$

provided that the expectation exists. The *standard deviation (sd) of $X$*, denoted by $\text{SD}(X)$, is defined by

$$\text{SD}(X) = +\sqrt{\text{Var}(X)}$$

- The mean/average of $(X-\mu)^2$, the squared deviation of $X$ from the expected value of $X$
- $\text{Var}(X) = 0 \Leftrightarrow X$ is a fixed constant equal to $\mu$
- $\text{Var}(X)$ has "strange" or usually meaningless units: e.g., when $X$ is in unit of "S\$", $\text{Var}(X)$ would be in unit of "S\$$^2$", while the sd of $X$ is in the same unit of "S\$" as $X$

# Deviation of Variance

Expand the LHS of the definition for variance,

$$
\begin{aligned}
E[(X-\mu)^2] &= E[X^2 - 2\mu X + \mu^2] \\
&= E[X^2] - 2\mu E[X] + \mu^2 \\
&= E[X^2] - 2\mu^2 + \mu^2 \\
&= E[X^2] - \mu^2
\end{aligned}
$$

- Proved by the linearity property of the expectation
- Stands for both discrete & cont. r.v.'s

**Computational Formula for Var($X$)**

For any r.v. $X$ with mean $\mu < \infty$, the *variance of $X$* can be equivalently computed as

$$\mathrm{Var}(X) = E(X^2) - \mu^2$$

provided that $E(X^2)$ exists, where

$$E(X^2) = \begin{cases} \sum_x x^2 p(x), & X \text{ is discrete with pmf } p(x) \\ \int x^2 f(x)\, dx, & X \text{ is cont. with pdf } f(x) \end{cases}$$

- It suffices to compute $E(X^2)$

# Variance & Standard Deviation III

**Variance of a Linear Transformation of a r.v.**

If $\text{Var}(X)$ exists & $Y = a + bX$ for some given constants $a, b \in \mathbb{R}$, then

$$\text{Var}(Y) = b^2 \text{Var}(X).$$

- It is intuitive that adding any constant $a$ to a r.v. $X$ does NOT affect the spread of a r.v./dist, as it merely shifts both the whole density $p/f$ & the center of the distribution, $E(X)$, by a units on the horizontal axis
- The <u>*sd of Y*</u> follows as <u>$SD(Y) = |b|SD(Y)$</u>

# Example: Variance & Standard Deviation

Consider the r.v. $X$ with pmf $p(x) = \begin{cases} .125, & x = 0, 3 \\ .375, & x = 1, 2 \\ 0, & \text{otherwise} \end{cases}$. Find $\mathrm{Var}(X)$

**Solution:**

- With $E(X) = 1.5$, by definition,

$$\begin{aligned}\mathrm{Var}(X) &= \sum_{x:p(x)>0} (x - E(X))^2 p(x) \\ &= (0 - 1.5)^2(.125) + (1 - 1.5)^2(.375) + (2 - 1.5)^2(.375) \\ &\quad + (3 - 1.5)^2(.125) \\ &= 2.25(.125) + .25(.375) + .25(.375) + 2.25(.125) = .75\end{aligned}$$

- Alternatively, by its computational formula & $E(X^2) = 3$,

$$\mathrm{Var}(X) = 3 - E(X)^2 = 3 - 1.5^2 = .75$$

- The sd of $X$ is given by $+\sqrt{.75} = .866$

# Example: Variance of a Bin(n, p) r.v.

Consider $X \sim Bin(n, p)$ with $p(x) = \binom{n}{x} p^x q^{n-x}$, $x = 0, 1, \ldots, n$. As $E(X^2) = E[X(X-1)] + E(X)$, it suffices to compute

$$E[X(X-1)] = \sum_{x=0}^{n} x(x-1)p(x) = \sum_{x=2}^{n} [x(x-1)] \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$$= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} q^{n-x} \quad [\text{let } y = x-2]$$

$$= n(n-1)p^2 \left[ \sum_{y=0}^{n-2} \frac{(n-2)!}{y!(n-2-y)!} p^y q^{n-2-y} \right] = n(n-1)p^2$$

Then, $Var(X) = n(n-1)p^2 + np - n^2p^2 = npq$

*Remark*: Of course, for $Y \sim Ber(p)$, one can work out $Var(Y) = pq$ either from definition or by setting $n = 1$ in this result

# Example: Variance of a B(a, b) r.v.

Consider $X \sim B(a, b)$. Compute

$$
\begin{aligned}
E(X^2) &= \int_0^1 (x^2) f(x)\, dx = C \int_0^1 (x^2) x^{a-1} (1-x)^{b-1}\, dx \\
&= C \int_0^1 x^{(a+2)-1} (1-x)^{b-1}\, dx = C \frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)} \\
&= \frac{a(a+1)}{(a+b)(a+b+1)}
\end{aligned}
$$

where $C = \Gamma(a+b)/[\Gamma(a)\Gamma(b)]$. Then,

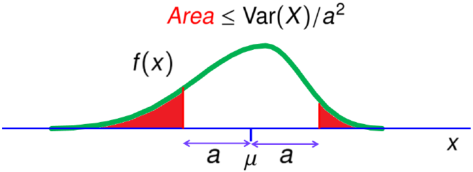
$$
Var(X) = \frac{a(a+1)}{(a+b)(a+b+1)} - \left(\frac{a}{a+b}\right)^2 = \frac{ab}{(a+b)^2(a+b+1)}
$$

knowing that $E(X) = \dfrac{a}{a+b}$

## Chebyshev's Inequality

For any r.v. $X$ with mean $\mu < \infty$, & $a > 0$, we have

$$P(|X - \mu| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}$$




- The prob that $X$ deviates much from its mean $\mu$ is low if $\sigma^2$ is very small; spread parameter $\sigma$
- Set $a = k\sigma \Rightarrow$ $P(|X - \mu| \geq k\sigma) \leq 1/k^2$
  - e.g., prob that $X \geq 4\sigma$ away from $\mu$ must be $\leq 1/16$
  - Standardized r.v.

- Either mean or variance provides a number which describes certain characteristics of a r.v./distribution. When it comes to $\geq 2$ r.v.'s, they can only be used as a criterion of comparison, but not of describing any relationship between the r.v.'s
- Define the *covariance* of 2 r.v.'s as a measure of their degree of *linear association* – degree to which $X$ & $Y$ "go together"

**How strong is the relationship/association between 2 r.v.'s?**

**Definition**

If $X$ & $Y$ are jointly distributed r.v.'s with finite marginal means $\mu_X$ & $\mu_Y$, respectively, the *covariance of $X$ & $Y$* is

$$\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X) = E[(X - \mu_X)(Y - \mu_Y)]$$

provided that the expectation exists.

# Covariance II

- Defined as an average of all the product of the deviations of $X$ from its mean & the deviation of $Y$ from its mean
- Computed by the result of $E[g(X,Y)]$
- **Note**: $\text{Var}(X) = \text{Cov}(X, X)$

### Computational Formula For Cov($X$, $Y$)

The covariance of $X$ & $Y$ can be equivalently computed by

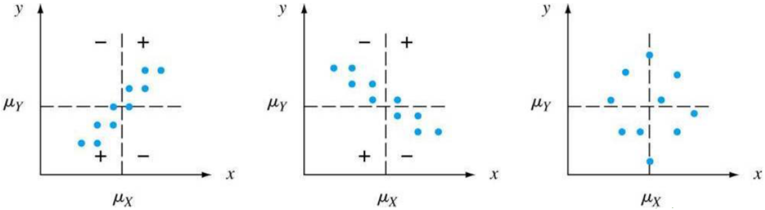
$$\text{Cov}(X,Y) = E(XY) - \mu_X \mu_Y$$

### Independence ⇒ Zero Covariance

Indep of $X$ & $Y$ implies zero covariance of $X$ & $Y$, but the converse is NOT true,

$$E(XY) = \mu_X \mu_Y \Rightarrow \text{Cov}(X,Y) = 0$$

- Ranges from $-\infty$ to $\infty$:
  - positive ($X$ & $Y$ are *positively correlated*): In general, both $X$ & $Y$ are larger/smaller than their respective means
  - negative ($X$ & $Y$ are *negatively correlated*): In general, one of $X$ & $Y$ is larger (resp. smaller) than its mean while the other is smaller (resp. larger) than its mean
  - zero ($X$ & $Y$ are *uncorrelated*): there are positive & negative products of deivations which offset each other
- Observe pairs of observations $(x, y)$ from the joint density: $\mathrm{Cov}(X, Y)$ is positive negative $\approx 0$



This is to illustrate an *important note about usage of covariance*:

**no casual effect to be concluded from non-zero covariance!**

It is well-known that *smoking causes lung cancer*. It is often observed that people who drink tend to have lung cancer (*i.e.,* drinking & having lung cancer are somehow “correlated”). However, it doe NOT mean that drinking causes lung cancer! It is because people who smoke usually also drink (*i.e.,* drinking & smoking are somehow “related”). In fact, any pair of drinking, smoking & tendency to have lung cancer are all *positively correlated*

# Example: Zero Covariance $\not\Rightarrow$ Independence

Here is a simple example illustrating the zero covariance implies indep. is NOT TRUE. Two dependent r.v.'s $X$ & $Y$ having zero covariance can be obtained by letting $X$ be a r.v. s.t.

$$P(X = 0) = P(X = 1) = P(X = -1) = \frac{1}{3}.$$

and define

$$Y = \begin{cases} 0, & X \neq 0 \\ 1, & X = 0. \end{cases}$$

Now, $XY = 0$ with prob 1, so $E(XY) = 0$. Also, $E(X) = 0$ & thus

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$$

We have zero covariance of $X$ & $Y$; however, $X$ & $Y$ are clearly not indept from the definition of $Y$.

# Example: Covariance of 2 Discrete r.v.'s

Suppose that the joint & the marginal pmf's for $X$ = "automobile policy deductible amount" & $Y$ = "homeowner policy deductible amount" are

| $p(x,y)$ | $y$ | | |
|---|---|---|---|
| $x$ | 0 | 100 | 200 |
| 100 | .20 | .10 | .20 |
| 250 | .05 | .15 | .30 |

| $x$ | 100 | 250 |
|---|---|---|
| $p_X(x)$ | .5 | .5 |

| $y$ | 0 | 100 | 200 |
|---|---|---|---|
| $p_Y(y)$ | .25 | .25 | .5 |

from which $\mu_X = 175$ & $\mu_Y = 125$. Then,

$$\begin{aligned}
\text{Cov}(X,Y) &= E(XY) - (175)(125) \\
&= (100)(100)(.1) + (100)(200)(.2) + (250)(100)(.15) \\
&\quad + (250)(200)(.3) - 21,875 \\
&= 1,875
\end{aligned}$$

Consider a random vector $(X, Y)$ that has a bivariate normal distribution with $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$, & $-1 < \rho < 1$. What is $\mathrm{Cov}(X, Y)$?

**Solution:** Note that $X, Y \sim N(0, 1)$. So, $E(X) = E(Y) = 0$, &

$$\mathrm{Cov}(X, Y) = E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{(xy)}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{\left(x^2 + y^2 - 2\rho xy\right)}{2(1-\rho^2)}\right] dy\, dx$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (xy) \exp\left[-\frac{(y-\rho x)^2 + x^2(1-\rho^2)}{2(1-\rho^2)}\right] dy\, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \left\{\int_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left[-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right] dy\right\} e^{-\frac{x^2}{2}}\, dx$$

$$= \int_{-\infty}^{\infty} x(\rho x) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = \rho$$

where the last equality follows as the integral is equivalent to $E(X^2) = \mathrm{Var}(X) = 1$, & the 2nd last equality follows as the inner integral (in $y$) equals $E(Y^*) = \rho x$ for $Y^* \sim N(\rho x, 1-\rho^2)$

**Covariance of 2 Sums of r.v.'s**

Suppose that $U = a + \sum_{i=1}^{n} b_i X_i$ & $V = c + \sum_{j=1}^{m} d_j Y_j$ for fixed constants $a, b_1, \cdots, b_n, c, d_1, \cdots, d_m \in \mathbb{R}$. Then,

$$\mathrm{Cov}(U, V) = \sum_{i=1}^{n} \sum_{j=1}^{m} b_i d_j \mathrm{Cov}(X_i, Y_j).$$

Some special cases of the above result about covariances:

- $\mathrm{Cov}(aX, bY) = ab\mathrm{Cov}(X, Y)$ for any $a, b \in \mathbb{R}$
- $\mathrm{Cov}(X + Y, Z + W) = \mathrm{Cov}(X, Z) + \mathrm{Cov}(X, W) + \mathrm{Cov}(Y, Z) + \mathrm{Cov}(Y, W)$

## Variance of a Linear Combination of r.v.'s

For any $n$ r.v.'s, $X_1, \cdots, X_n$, & fixed constants $a_1, \cdots, a_n, b_1, \cdots, b_n$,

$$\mathrm{Var}(a + \sum_{i=1}^{n} b_i X_i) = \sum_{i=1}^{n} b_i^2 \mathrm{Var}(X_i) + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} b_i b_j \mathrm{Cov}(X_i, X_j)$$

A special case of the above result about covariances:

- $\mathrm{Var}(X \pm Y) = \mathrm{Cov}(X \pm Y, X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\mathrm{Cov}(X, Y)$

## Variance of a Sum of Independent r.v.'s

Suppose that $X_1, \cdots, X_n$ are $n$ indept r.v.'s. Then,

$$\mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

**Definition**

If $X$ & $Y$ are jointly distributed r.v.'s with finite marginal means $\mu_X$ & $\mu_Y$, respectively, the *correlation coefficient of $X$ & $Y$* is

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

provided that the covariance and variances exist.

- $\rho(X,Y) = \rho(Y,X)$
- By the way $\text{Corr}(X,Y)$ is formed, $\text{Corr}(X,Y)$ is a dimension-less quantity, i.e., $\text{Corr}(X,Y)$ has no units
- For any constants $a, b, c, d$ and r.v.'s $X, Y$, $\text{Corr}(aX+b, cY+d) = \text{Corr}(X,Y)$ $\Rightarrow$ correlation coefficient is invariant under the linear transformation of two r.v.'s

# Correlation Coefficient II

**Properties of Correlation Coefficient**

If $X$ & $Y$ are jointly distributed r.v.'s with correlation coefficient $\rho(X, Y)$, then

$$-1 \leq \rho(X, Y) \leq Y.$$

Furthermore, $\rho(X, Y) = \pm 1$ if and only if $P(Y = a + bX) = 1$ for some constants $a$ and $b$.

- Proof.

$$\begin{aligned} 0 &\leq \mathrm{Var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) = \mathrm{Var}\left(\frac{X}{\sigma_X}\right) + \mathrm{Var}\left(\frac{Y}{\sigma_Y}\right) + 2\mathrm{Cov}\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right) \\ &= \frac{\mathrm{Var}(X)}{\sigma_X^2} + \frac{\mathrm{Var}(Y)}{\sigma_Y^2} + \frac{2\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} \\ &= 2(1 + \rho(X, Y)) \end{aligned}$$

  So, $\rho(X, Y) \geq -1$. Similarly, $\mathrm{Var}(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}) \geq 0 \Rightarrow \rho(X, Y) \leq 1$.

- When $\rho = \pm 1$, $Y$ can be viewed as a linear transformation of $X$
- A more useful measure of relationship/association between 2 r.v.'s

# Example: Correlation Coefficient

- Example on Page 40.

$$\begin{aligned}\mathrm{Var}(X) &= E(X^2) - [E(X)]^2 \\ &= 100^2(.5) + (250^2)(.5) - (100 * .5 + 250 * .5)^2 = 5625 \\ \mathrm{Var}(Y) &= E(Y^2) - [E(Y)]^2 \\ &= 0^2(.25) + 100^2(.25) + 200^2(.5) - (0(.25) + 100(.25) + 200(.5))^2 \\ &= 6875\end{aligned}$$

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{1875}{\sqrt{5625 * 6875}} = .3015$$

- Bivariate normal vector.
  Refer to Example on Page 41,

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\rho}{1 \cdot 1} = \rho.$$

  For any bivariate normal distribution with $\sigma_X$, $\sigma_Y$, the correlation coefficient is also $\rho$

# Conditional Expectation I

- We discuss how to find the conditional distribution $X|Y=y$ in Lecture 3, which can be viewed as a new r.v.

**How to find the expectation of this conditional dist?**

## Conditional Expectation

Suppose that the conditional distribution of $X|Y=y$ is known.

- If $X$ and $Y$ are discrete r.v.'s, and the conditional pmf is $p_{X|Y}(x|y)$, then the *conditional expectation* for $X|Y=y$ is

$$E(X|Y=y) = \sum_x x p_{X|Y}(x|y)$$

- If $X$ and $Y$ are cont. r.v.'s, and the conditional pdf is $f_{X|Y}(x|y)$, then the *conditional expectation* for $X|Y=y$ is

$$E(X|Y=y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx$$

# Example: Conditional Expectation

- Example on Page 41. Interest in $X|Y = 0$. Find the conditional pmf,

$$p_{X|Y}(100|0) = P(X = 100|Y = 0) = \frac{P(X = 100, Y = 0)}{P(Y = 0)} = \frac{.20}{.25} = .8$$

$$p_{X|Y}(250|0) = P(X = 250|Y = 0) = \frac{P(X = 250, Y = 0)}{P(Y = 0)} = \frac{.05}{.25} = .2$$

$p_{X|Y}(x|0) = 0$ for $x \neq 250$ and $x \neq 100$. The conditional expectation is

$$E[X|Y = 0] = 100(.8) + 250(.2) = 130.$$

- Consider the bivariate normal vector. Recall that in Lecture 3, we find

$$X|Y = y \sim N(\mu_X + \rho\sigma_X z_y, (1 - \rho^2)\sigma_X^2),$$

where $z_y = (y - \mu_Y)/\sigma_Y$. So, $E(X|Y = y) = \mu_X + \rho(y - \mu_Y)\sigma_X/\sigma_Y$

# Conditional Expectation II

- $X|Y = y$ is a new r.v., we apply the definition of expectation to this r.v.
- Assume the conditional expectation of $X$ given $Y = y$ exists for every $y$ in the range of $Y$, hence $E(X|Y)$ is a new random variable. It can be viewed as a function of $Y$, where $g(y) = E(X|Y = y)$.
- e.g. for bivariate normal vector,

$$E(X|Y) = \mu_X + \rho(Y - \mu_Y)\sigma_X/\sigma_Y.$$

  Because $Y \sim N(\mu_Y, \sigma_Y^2)$, and this is a linear transformation of $Y$, so $E(X|Y) \sim N(\mu_X, \rho^2\sigma_X^2)$ is a normal r.v.
- Generally, $E(X|Y)$ is different from $Y$, or $X$.

# Law of Total Expectation

**Law of Total Expectation**

For two r.v.'s $X$ and $Y$, if the expectation and conditional expectation exist,

$$E(X) = E[E(X|Y)].$$

- Proof for discrete case:

$$\begin{aligned} \text{RHS} &= \sum_y E(X|Y=y)p_Y(y) = \sum_y p_Y(y)(\sum_x p_{X|Y}(x|y)x) \\ &= \sum_y \sum_x x p_{X|Y}(x|y)p_Y(y) \\ &= \sum_x x[\sum_y p_{X|Y}(x|y)p_Y(y)] = \sum_x x p_X(x) = E(X) \end{aligned}$$

- The expectation of a r.v. $X$ can be calculated by calculating the conditional expectations first, and then summing/integrating the weighted cond. expectations

# Example: Law of Total Expectation

Suppose in a system, a component and a backup unit both have mean lifetimes $\mu = 5$ years. If the component fails, the system automatically substitutes the backup unit, but here is probability $p = .1$ that something will go wrong and it will fail to substitute. What is the expectation for the lifetime of this system?

**Solution** . Let $T$ be the total lifetime, and let $X = 1$ if the substitution of the backup is successful, & $X = 0$ if it fails. The total lifetime is the lifetime of the component only if $X = 0$, and the sum of the lifetimes of the original and backup units if $X = 1$.

$$E(T|X = 1) = 10, \qquad E(T|X = 0) = 5$$

Thus, the expectation of total lifetime is

$$\begin{aligned} E(T) &= E(T|X = 1)P(X = 1) + E(T|X = 0)P(X = 0) \\ &= 10(.9) + 5(.1) = 9.5 \text{ years} \end{aligned}$$

# Example: Random Sum

On the day before the exam, each student entering the TA's office will ask one question that will come out for the exam with probability $p = .05$. The number of student going to office hours that day is Poisson distributed with parameter $\lambda = 10$. In the exam, what is the expected number of questions that have been asked by students?

**Solution**. Let $N$ be the number of students coming to office hours that day, then $N \sim Pois(10)$. Let $X_i$ be the number of questions that will appear in the exam asked by $i$th student, then $X_i \sim Ber(.05)$, any integer $i$. Let $X = \sum_{i=1}^{N} X_i$, the expectation of total expected number of questions is $E(X) = E[E(X|N)]$. Since

$$E(X|N=n) = E[\sum_{i=1}^{N} X_i | N = n] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = nE[X_1].$$

According to law of total expectation,

$$E(X) = E[E(X|N)] = E[NE[X_1]] = E[N]E[N_1] = 10(.05) = .5.$$

**Definition**

The moment generating funciton (mgf) of a r.v. $X$ is defined by, for a constant $t \in \mathbb{R}$,

$$M(t) = E(e^{tX}) = \begin{cases} \sum_x e^{tx} p(x), & \text{when } X \text{ is discrete} \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx, & \text{when } X \text{ is cont.} \end{cases}$$

if the expectation exists.

- $M(t)$ is a function of $t$ (i.e., does not contain $x$ anymore)
- Generates all the *moments* $E(X^k)$, for $k = 1, 2, \cdots$, of the r.v. $X$
- Due to the usage of $M(t)$, we only care about $t$ to be in an open interval containing zero $\Leftrightarrow t \in (-a, b)$ for some $a, b \in \mathbb{R}$

# Moment Generating Function of a r.v. II

**Characterization of a r.v.**

If the mgf of $X$ exists for $t$ in an open interval containing zero, it *uniquely determines the probability distribution of $X$*.

- In finding the distribution of a r.v., we can *find its mgf* (apart from pdf/pmf or cdf) & then deduce its distribution by matching against a list of mgf's for some standard & common probability distributions

**Determination of the $k$th moment of a r.v.**

If the mgf of $X$ exists for $t$ in an open interval containing zero, then the *$k$th moment of $X$* is given by

$$E(X^k) = M^{(k)}(0) = \frac{d^k}{dt^k} M(t)|_{t=0}$$

# Example: mgf & Moments of a Bin(n, p) r.v.

For $X \sim Bin(n,p)$, its mgf $M(t)$ equals

$$E(e^{tX}) = \sum_x e^{tx} p(x) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x} = (pe^t + q)^n$$

Now, we can obtain some moments & variance of this r.v.:

- Differentiating $M(t)$ once yields

$$M'(t) = n(pe^t + q)^{n-1} pe^t, \qquad \& \qquad E(X) = M'(0) = np$$

- Differentiating $M(t)$ twice yields

$$M^{(2)}(t) = n(n-1)(pe^t + q)^{n-2}(pe^t)^2 + n(pe^t + q)^{n-1} pe^t$$

& $E(X^2) = M^{(2)}(0) = n(n-1)p^2 + np$
Hence, $\text{Var}(X) = E(X^2) - [E(X)]^2 = npq$

# Example: mgf of Some Discrete r.v.'s

| | Probability mass function, $p(x)$ | Moment generating function, $M(t)$ | Mean | Variance |
|---|---|---|---|---|
| Binomial with parameters $n, p$; $0 \le p \le 1$ | $\binom{n}{x} p^x(1-p)^{n-x}$ <br> $x = 0, 1, \ldots, n$ | $(pe^t + 1 - p)^n$ | $np$ | $np(1-p)$ |
| Poisson with parameter $\lambda > 0$ | $e^{-\lambda}\frac{\lambda^x}{x!}$ <br> $x = 0, 1, 2, \ldots$ | $\exp\{\lambda(e^t - 1)\}$ | $\lambda$ | $\lambda$ |
| Geometric with parameter $0 \le p \le 1$ | $p(1-p)^{x-1}$ <br> $x = 1, 2, \ldots$ | $\frac{pe^t}{1-(1-p)e^t}$ | $\frac{1}{p}$ | $\frac{1-p}{p^2}$ |
| Negative binomial with parameters $r, p$; $0 \le p \le 1$ | $\binom{n-1}{r-1} p^r(1-p)^{n-r}$ <br> $n = r, r+1, \ldots$ | $\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$ | $\frac{r}{p}$ | $\frac{r(1-p)}{p^2}$ |

- For $X \sim U(a,b)$, its mgf $M(t)$ equals

$$E(e^{tX}) = \int_a^b \frac{e^{tx}}{b-1}dx = \frac{1}{b-a}\int_a^b e^{tx}dx = \frac{1}{b-a}\frac{e^{tx}}{t}\Big|_a^b = \frac{e^{bt}-e^{at}}{(b-a)t}$$

- For $X \sim N(\mu, \sigma^2)$, its mgf $M(t)$ equals

$$\begin{aligned}
E(e^{tX}) &= \int_{-\infty}^{\infty} e^{tx}\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/(2\sigma^2)}dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}e^{-[x^2+\mu^2-2(\mu+\sigma^2 t)x]/(2\sigma^2)}dx \\
&= e^{-\mu^2/(2\sigma^2)}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}e^{-[x^2-2(\mu+\sigma^2 t)x]/(2\sigma^2)}dx \\
&= e^{-\mu^2/(2\sigma^2)}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}\{[x-(\mu+\sigma^2 t)]^2-(\mu+\sigma^2 t)^2\}}dx \\
&= e^{-\frac{1}{2\sigma^2}[-\mu^2+(\mu+\sigma^2 t)^2]}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}[x-(\mu+\sigma^2 t)]^2}dx \\
&= e^{\mu t}e^{\sigma^2 t^2/2}
\end{aligned}$$

# Example: mgf of Some Cont. r.v.'s

| | Probability density function , $f(x)$ | Moment generating function, $M(t)$ | Mean | Variance |
|---|---|---|---|---|
| Uniform over $(a, b)$ | $f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$ | $\frac{e^{tb} - e^{ta}}{t(b-a)}$ | $\frac{a+b}{2}$ | $\frac{(b-a)^2}{12}$ |
| Exponential with parameter $\lambda > 0$ | $f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$ | $\frac{\lambda}{\lambda - t}$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ |
| Gamma with parameters $(s, \lambda), \lambda > 0$ | $f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{s-1}}{\Gamma(s)} & x \geq 0 \\ 0 & x < 0 \end{cases}$ | $\left(\frac{\lambda}{\lambda - t}\right)^s$ | $\frac{s}{\lambda}$ | $\frac{s}{\lambda^2}$ |
| Normal with parameters $(\mu, \sigma^2)$ | $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$ $-\infty < x < \infty$ | $\exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$ | $\mu$ | $\sigma^2$ |

# MGF of a Function of a r.v./r.v.'s

**mgf of a Linear Transformation of a r.v.**

If $X$ has the mgf $M_X(t)$, for constants $a, b \in \mathbb{R}$, the *mgf of $Y = a + bX$* equals

$$M_Y(t) = e^{at} M_X(bt)$$

**mgf of a Sum of Independent r.v.'s**

If $X_1, \ldots, X_n$ are indept r.v.'s with mgf's $M_{X_i}$, then the *mgf of $Z = X_1 + X_2 + \cdots + X_n$*

$$M_Z(t) = M_{X_1}(t) \times M_{X_2}(t) \times \cdots \times M_{X_n}(t)$$

on the common interval where all the $n$ mgf's at the RHS exist.

- Both of the above results are extremely useful as they require only mgf's of individual r.v.'s but not manipulation of any joint densities

Note that the mgf of $X \sim U(a,b)$ is

$$M_X(t) = \frac{e^{bt} - e^{at}}{(b-a)t}$$

Obviously, setting $a = 0$ and $b = 1$ yields the mgf of $U \sim U(0,1)$ as

$$M_U(t) = \frac{e^t - 1}{t}$$

Re-arrange the above expression of $M_X(t)$ as in a linear combination form:

$$M_X(t) = e^{at}\frac{e^{(b-a)t} - 1}{(b-a)t} = e^{at}M_U((b-a)t)$$

Hence, $X \sim U(a,b)$ is a linear transformation of $U \sim U(0,1)$ through $X = a + (b-a)U$

# Example: Sum of Indept Poisson r.v.'s

- For Poisson r.v., we have proved that when $X$ & $Y$ are indept Poisson r.v.'s with parameters $\lambda > 0$ & $\mu > 0$ respectively, the sum $Z = X + Y$ would be a $Poi(\lambda + \mu)$ r.v. using the convolution formula via manipulating the 2 Poisson pmf's
- In fact, such a result can be shown using mgf's without much effort

First of all, the mgf of a $Poi(\lambda)$ r.v. $X$ is given from the table as

$$M_X(t) = \exp(\lambda(e^t - 1))$$

By mgf of a sum of indept r.v.'s, the mgf of $Z = X + Y$ is given by

$$M_Z(t) = M_X(t) \times M_Y(t) = \exp(\lambda(e^t - 1)) \exp(\mu(e^t - 1)) = e^{(\lambda+\mu)(e^t - 1)}.$$

Checking this mgf against that of a Poisson r.v. concludes that $Z = X + Y \sim Poi(\lambda + \mu)$ as $\lambda + \mu > 0$.