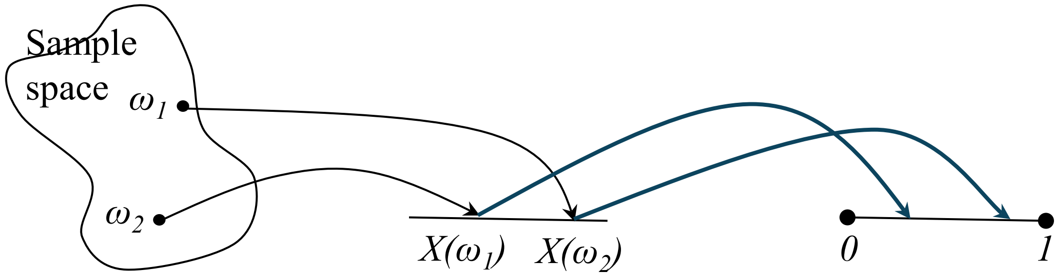
# Random Variable

# Random Variable Definition

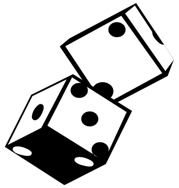
- It frequently occurs that in performing an experiment we are mainly interested in some functions of the outcome as opposed to the outcome itself
  - Rotery, casino, game, etc.
- These quantities of interest, or more formally, these real-valued functions defined on the sample space, are known as random variables (RVs).
- More mathematical expression of RVs
  - Given a probability space $(S, F, P)$, a random variable is a measurable function (mapping) from $S$ to the real line $X$: $S \rightarrow R$
    - $X(\{H\}) = 100, X(\{H\}) = -50$
    - $X(\{1,2\}) = 100, X(\{3,4,5,6\}) = -50$
- Since the value of a random variable is defined by the outcome of the experiment, we may assign probabilities to the possible values of the random variable

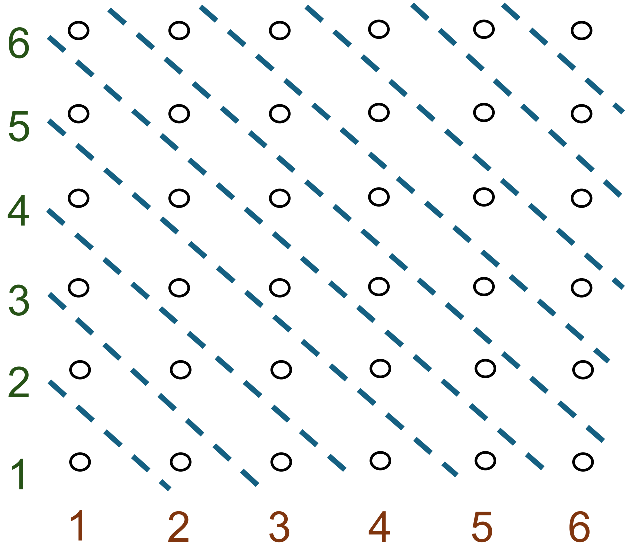


# Random Variable: Examples

- Ex 2.1] Letting $X$ denote the random variable that is defined as the sum of two fair dice
  - The outcomes of two dice = $(\omega_1, \omega_2)$
  - RV $X(\omega_1, \omega_2) = \omega_1 + \omega_2$
  - Possible values of $X = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$
  - Probabilities to the possible values of the random variable
    - $P(X = 2) = P\{(1,1)\} = 1/36$
    - $P(X = 3) = P\{(1,2), (2,1)\} = 2/36$

    ...
    - $P(X = 11) = P\{(5,6), (6,5)\} = 2/36$
    - $P(X = 12) = P\{(6,6)\} = 1/36$
    - $1 = P\{\cup_{n=2}^{12} \{X = n\}\} = \sum_{n=2}^{12} P(X = n)$





# Random Variable: Examples

- Ex.2.3] Suppose that we toss a coin having a probability $p$ of coming up heads, until the first head appears. Letting $N$ denote the number of flips required, then assuming that the outcome of successive flips are independent, $N$ is a random variable taking on one of the values 1, 2, 3, . . . , with respective probabilities

$$P\{N = 1\} = P\{H\} = p,$$
$$P\{N = 2\} = P\{(T, H)\} = (1 - p)p,$$
$$P\{N = 3\} = P\{(T, T, H)\} = (1 - p)2p,$$
$$\dots$$
$$P\{N = n\} = P\{(T, T, \dots, T, H)\} = (1 - p)^{n-1}p, n \geq 1$$

- Indicator random variable

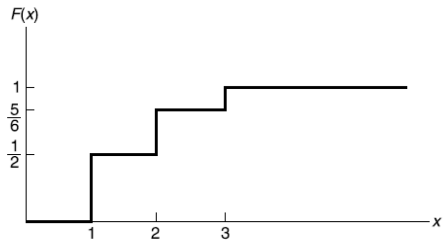
  $I_E(x) = 1$ if $x \in E$, 0 otherwise.

# Cumulative Distribution Function (CDF)

- If a random variable takes on either a finite or a countable number of possible values, the RV is called discrete.
- If a continuum of possible values, continuous.
- The cumulative distribution function (CDF) $F(\ )$ of the random variable X is defined for any real number $b$, $-\infty < b < \infty$ by $F(b) = P\{X \leq b\}$.
- Properties of CDF F
  1. $F(b)$ is a nondecreasing function of b
  2. $lim_{b \to \infty} F(b) = F(\infty) = 1$
  3. $lim_{b \to -\infty} F(b) = F(-\infty) = 0$
- Ex. $P\{a < X \leq b\} = F(b) - F(a)$ for all $a < b$

# Discrete RVs

- A random variable that can take on at most countable number of possible values
- Probability Mass Function (pmf): for a discrete random variable $X$, pmf $p(a)$ of $X$ is defined a $p(a) = P\{X = a\}$.
- Properties of pmf
  1. $p(x_i) > 0, i = 1,2, \ldots$
  2. $p(x) = 0$ all other values of $x$
  3. $\Sigma_{i=\{1,\ldots,\infty\}} p(x_i) = 1$
  4. cdf vs. pmf: $F(a) = \sum_{x_i \le a} p(xi)$
- Ex. $p(1) = 1/2, p(2) = 1/3, p(3) = 1/6$
  $F(a) = 0 \quad a < 1, F(a) = \frac{1}{2} \; 1 \le a < 2, F(a) = \frac{5}{6} \; 2 \le a < 3, F(a) = 1 \; 3 \le a$




- Discrete random variables are often classified according to their pmf
  - Ex: Bernoulli RV, Binomial RV, Geometric RV, Poisson RV

# Example of Discrete RVs: The Bernoulli RV

- A random variable X is said to be a Bernoulli RV for given sample space $S = \{A, A^c\}$ and some $p \in (0,1)$ if its pmf is given by
$$P\{A\} = p, P\{A^c\} = 1 - p$$

- For example, suppose that a trial, or an experiment, whose outcome can be classified as either a "success" or as a "failure" is performed. If we let $X$ equal 1 if the outcome is a success and 0 if it is a failure, then the probability mass function of X is given by
$$p(0) = P\{X = 0\} = 1 - p,$$
$$p(1) = P\{X = 1\} = p \qquad (2.2)$$
where $p, 0 \leq p \leq 1$, is the probability that the trial is a "success."

# Example of Discrete RVs: The Binomial RV

- Suppose that n independent trials, each of which results in a “success” with probability p and in a “failure” with probability $1 - p$, are to be performed.

- If X represents the number of successes that occur in the $n$ trials, then $X$ is said to be a binomial random variable with parameters $(n, p)$.

- The pmf of a binomial RV with (n,p)
  $p(i) = \binom{n}{i} p^i (1-p)^{n-i}$ for $i \ = \ 0, 1, \ldots, n$ (2.3)
  where $\binom{n}{i} = \frac{n!}{(n-i)!i!}$.

- $\sum_{i=0}^{n} p(i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = (p + (1-p))^n = 1$

# Example of Discrete RVs: The Binomial RV

- Ex. 2.7] It is known that any item produced by a certain machine will be defective with probability 0.1, independently of any other item. What is the probability that in a sample of three items, at most one will be defective?
  - If $X$ is the number of defective items in the sample, then $X$ is a binomial random variable with parameters $(3, 0.1)$. Hence, the desired probability is given by
  - $P\{X = 0\} + P\{X = 1\} \ = (3,0)(0.1)0(0.9)3 + (3,1)(0.1)1(0.9)2 = 0.972$

# Example of Discrete RVs: The Geometric RV

- Suppose that independent trials, each having probability p of being a success, are performed until a success occurs.

- If we let $X$ be the number of trials required until the first success, then $X$ is said to be a geometric random variable with parameter $p$.

- Its probability mass function is given by
$p(n) = P\{X = n\} = (1 - p)^{n-1}p, n = 1, 2, \ldots$ (2.4)

- To check that p(n) is a probability mass function, we note that
$\sum_{n=1}^{\infty} p(n) = p \sum_{n=1}^{\infty} (1 - p)^{n-1} = 1$

# Example of Discrete RVs: The Poisson RV

- A random variable $X$, taking on one of the values $0, 1, 2, \ldots$, is said to be a Poisson random variable with parameter $\lambda$, if for some $\lambda > 0$,
  $p(i) = P\{X = i\} = e^{-\lambda}\frac{\lambda^i}{i!}, i = 0,1,2,3, \ldots$ (2.5)

- To check that p(n) is a probability mass function, we note that
  $\sum_{i=0}^{\infty} p(i) = \sum_{i=0}^{\infty} e^{-\lambda}\frac{\lambda^i}{i!} = e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} = e^{-\lambda}e^{\lambda} = 1$

- Binomial vs. Poisson RVs
  - $n \gg 1, p \ll 1$, let $\lambda = np$
  - $P\{X = 1\} = \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{(n-i)!i!}\left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} = \frac{n(n-1)\ldots(n-i+1)}{n^i}\frac{\lambda^i}{i!}\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^i}$
  - For large n and small p
    - $\left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda}, \left(1 - \frac{\lambda}{n}\right)^i \approx 1, \frac{n(n-1)\ldots(n-i+1)}{n^i} \approx 1$
  - $P\{X = i\} \approx e^{-\lambda}\frac{\lambda^i}{i!}$

# Example of Discrete RVs: The Poisson RV

- Ex. 2.10] Suppose that the number of typographical errors on a single page of this book has a Poisson distribution with parameter $\lambda = 1$. Calculate the probability that there is at least one error on this page.
  - $P\{X \geq 1\} = 1 - P\{X = 0\} = 1 - e^{-1} = 0.633$

| Discrete probability distribution | Probability mass function, $p(x)$ | Moment generating function, $\phi(t)$ | Mean | Variance |
|---|---|---|---|---|
| Binomial with parameters $n, p$, $0 \leq p \leq 1$ | $\binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n$ | $(pe^t + (1-p))^n$ | $np$ | $np(1-p)$ |
| Poisson with parameter $\lambda > 0$ | $e^{-\lambda} \frac{\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$ | $\exp\{\lambda(e^t - 1)\}$ | $\lambda$ | $\lambda$ |
| Geometric with parameter $0 \leq p \leq 1$ | $p(1-p)^{x-1}$, $x = 1, 2, \ldots$ | $\frac{pe^t}{1-(1-p)e^t}$ | $\frac{1}{p}$ | $\frac{1-p}{p^2}$ |

# Continuous RVs

- RV X is said to be a continuous RV if there exists a non-negative function $f(x)$, defined for all real $x \in \{-\infty, \infty\}$ having the property that for any set B of real numbers

$$P\{X \in B\} = \int_B f(x)dx = 1$$

- The function $f(x)$ is called the probability density function (pdf) of the random variable $X$

- Properties of pdf of X
  - $P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)dx = 1$
  - $P\{a \leq X \leq b\} = \int_a^b f(x)dx$
  - $P\{X = a\} = \int_a^a f(x)dx = 0$
  - $F(a) = P\{X \in (-\infty, a)\} = \int_{-\infty}^{a} f(x)dx$
  - $\frac{dF(a)}{da} = f(a)$

- Ex. Uniform Rv, Exponential Rv, Gamma RV, Normal RV

# Example of Continuous RVs: Uniform RV

- A RV is said to be uniformly distributed over the interval $(\alpha, \beta)$ if its probability density function is given by

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha}, \alpha < x < \beta \\ 0, otherwise \end{cases}$$

- Ex. 2.14] $X \sim U(0,10)$
  - $P\{X < 3\} = \int_{-\infty}^{3} f(x)dx = \int_{0}^{3} \frac{1}{10} dx = \frac{3}{10}$
  - $P\{X > 7\} = \int_{7}^{\infty} f(x)dx = \int_{7}^{10} \frac{1}{10} dx = \frac{3}{10}$
  - $P\{1 \leq X < 6\} = \int_{1}^{6} f(x)dx = \int_{1}^{6} \frac{1}{10} dx = \frac{5}{10} = \frac{1}{2}$

# Example of Continuous RVs: Exponential RV

- A continuous RV is said to be an exponential RV with parameter $\lambda$ if its probability density function is given, for $\lambda > 0$, by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, if\ x \geq 0 \\ 0, otherwise \end{cases}$$

$$F(a) = \int_0^a \lambda e^{-\lambda x} dx = 1 - e^{-\lambda a}$$

# Example of Continuous RVs: Gamma RV

- A continuous RV is said to be a gamma RV with parameter $\alpha, \lambda$, if its probability density function is given, for $\alpha > 0, \lambda > 0$, by

$$f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, if\ x \geq 0 \\ 0, otherwise \end{cases}$$

where

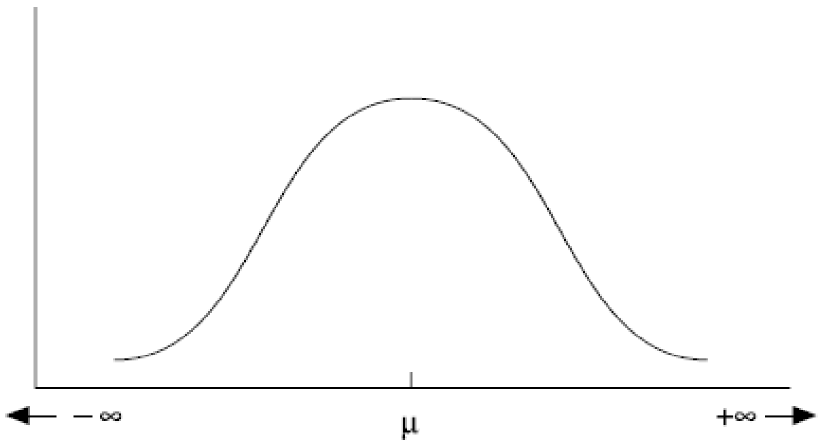
$$\Gamma(\alpha) = \int_0^{\infty} e^{-x} x^{\alpha-1} dx$$

# Example of Continuous RVs: Normal RV

- A continuous RV is said to be a normal RV with parameter $\mu$ and $\sigma^2$, if its probability density function is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

for $-\infty < x < \infty$.



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

| Discrete probability distribution | Probability mass function, $p(x)$ | Moment generating function, $\phi(t)$ | Mean | Variance |
|---|---|---|---|---|
| Binomial with parameters $n, p$, $0 \le p \le 1$ | $\binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n$ | $(pe^t + (1-p))^n$ | $np$ | $np(1-p)$ |
| Poisson with parameter $\lambda > 0$ | $e^{-\lambda} \frac{\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$ | $\exp\{\lambda(e^t - 1)\}$ | $\lambda$ | $\lambda$ |
| Geometric with parameter $0 \le p \le 1$ | $p(1-p)^{x-1}$, $x = 1, 2, \ldots$ | $\frac{pe^t}{1-(1-p)e^t}$ | $\frac{1}{p}$ | $\frac{1-p}{p^2}$ |

| Continuous probability distribution | Probability density function, $f(x)$ | Moment generating function, $\phi(t)$ | 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 $\lambda > 0$ | $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x < 0 \end{cases}$ | $\frac{\lambda}{\lambda - t}$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ |
| Gamma with parameters $(n, \lambda)$, $\lambda > 0$ | $f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!}, & x \ge 0 \\ 0, & x < 0 \end{cases}$ | $\left(\frac{\lambda}{\lambda - t}\right)^n$ | $\frac{n}{\lambda}$ | $\frac{n}{\lambda^2}$ |
| Normal with parameters $(\mu, \sigma^2)$ | $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \times \exp\{-(x-\mu)^2/2\sigma^2\}$, $-\infty < x < \infty$ | $\exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$ | $\mu$ | $\sigma^2$ |

# Expectation of a function of a RV

- The expected value of $X$ , $E[X]$, is a weighted average of the possible values that $X$ can take on, each value being weighted by the probability that $X$ assumes that value.

- If RV X is a discrete RV having a pmf $p(x)$, then $E[X]$ is defined by

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

  - Ex. 2.16~2.19] find the followings
    - Expectation of a Bernoulli Random Variable
    - Expectation of a Binomial Random Variable
    - Expectation of a Geometric Random Variable
    - Expectation of a Poisson Random Variable

# Expectation of a function of a RV

- If RV X is a continuous RV having a pdf $f(x)$, then $E[X]$ is defined by

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

  - Ex. 2.20~2.22] find the followings
    - Expectation of a Uniform Random Variable
    - Expectation of an Exponential Random Variable
    - Expectation of a Gamma Random Variable
    - Expectation of a Normal Random Variable

| Discrete probability distribution | Probability mass function, $p(x)$ | Moment generating function, $\phi(t)$ | Mean | Variance |
|---|---|---|---|---|
| Binomial with parameters $n, p$, $0 \le p \le 1$ | $\binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n$ | $(pe^t + (1-p))^n$ | $np$ | $np(1-p)$ |
| Poisson with parameter $\lambda > 0$ | $e^{-\lambda} \frac{\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$ | $\exp\{\lambda(e^t - 1)\}$ | $\lambda$ | $\lambda$ |
| Geometric with parameter $0 \le p \le 1$ | $p(1-p)^{x-1}$, $x = 1, 2, \ldots$ | $\frac{pe^t}{1-(1-p)e^t}$ | $\frac{1}{p}$ | $\frac{1-p}{p^2}$ |

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

# Expectation of a function of a RV

- Given a RV $X$ and its probability distribution, what is the expectation of a function of $X$?
  1. Since $g(X)$ is itself a random variable, it must have a probability distribution, which should be computable from a knowledge of the distribution of $X$. Once we have obtained the distribution of $g(X)$, we can then compute $E[g(X)]$ by the definition of the expectation.
  2. Another way is to compute the expectation of a function of $X$ from a knowledge of the distribution of $X$. See Proposition 2.1.

# Expectation of a function of a RV

- Ex. 2.23] Suppose $X$ has the following probability mass function: $p(0) = 0.2, p(1) = 0.5, p(2) = 0.3$. Calculate $E[X^2]$.
  - Letting $Y = X^2$, we have that $Y$ is a random variable that can take on one of the values $0^2, 1^2, 2^2$ with respective probabilities
$$p_Y(0) = P\{Y = 02\} = 0.2,$$
$$p_Y(1) = P\{Y = 12\} = 0.5,$$
$$p_Y(4) = P\{Y = 22\} = 0.3$$
  - Hence, $E[X^2] = E[Y] = 0(0.2) + 1(0.5) + 4(0.3) = 1.7$
  - Note that $E[X^2] \neq E[X]^2$

# Expectation of a function of a RV

- Proposition 2.1
  1. If $X$ is a discrete random variable with probability mass function $p(x)$, then for any real-valued function $g$,
  $$E[g(X)] = \sum_{x:p(x)>0} g(x)p(x)$$
  2. If $X$ is a continuous random variable with probability density function $f(x)$, then for any real-valued function $g$,
  $$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\,dx$$

- Ex 2.26] Let X be uniformly distributed over (0,1), $E[X^3]$?
  - $E[X^3] = \int_0^1 X^3\,dx = \frac{1}{4}$

- Corollary 2.2
  - If $a$ and $b$ are constants, then $E[aX + b] = aE[X] + b$

# Expectation of a function of a RV

- The expected value of a RV $X$, $E[X]$, is also referred to as the mean or the first moment of $X$.

- The quantity $E[X^n], n \geq 1$, is called the $n_{th}$ moment of $X$

- The variance of a RV $X$, denoted by $Var(X)$, is defined by $Var(X) = E[(X - E[X])^2]$, deviation of $X$ from the mean.

- $Var(X) = E[(X - E[X])^2]$
  $= E[X^2 - 2E[X]X + E[X]^2]$
  $= E[X^2] - E[2E[X]X] + E[E[X]^2]$
  $= E[X^2] - 2E[X]^2 + E[X]^2$
  $= E[X^2] - E[X]^2$

# Expectation of a function of a RV

- Ex 2.27] $Var(X)$ of the normal RV with $\mu$ and $\sigma$.
  - $Var(X) = E[(X-\mu)^2]$

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

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-\frac{y^2}{2}} dy$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy$$

$$= \sigma^2$$

| Discrete probability distribution | Probability mass function, $p(x)$ | Moment generating function, $\phi(t)$ | Mean | Variance |
|---|---|---|---|---|
| Binomial with parameters $n, p$, $0 \leq p \leq 1$ | $\binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n$ | $(pe^t + (1-p))^n$ | $np$ | $np(1-p)$ |
| Poisson with parameter $\lambda > 0$ | $e^{-\lambda} \frac{\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$ | $\exp\{\lambda(e^t - 1)\}$ | $\lambda$ | $\lambda$ |
| Geometric with parameter $0 \leq p \leq 1$ | $p(1-p)^{x-1}$, $x = 1, 2, \ldots$ | $\frac{pe^t}{1-(1-p)e^t}$ | $\frac{1}{p}$ | $\frac{1-p}{p^2}$ |

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

# Jointly Distributed RVs

- For any two RVs $X$ and $Y$, the joint cumulative probability distribution function of $X$ and $Y$ is defined as $F(a, b) = P\{X \leq a, Y \leq b\}, -\infty \leq a, b \leq \infty$
  - $F_X(a, b) = P\{X \leq a\} = P\{X \leq a, Y \leq \infty\} = F(a, \infty)$
  - If $X$ and $Y$ are discrete RVs, the joint pmf of $X$ and $Y$ is defined as $p(x, y) = P\{X = x, Y = y\}$
- The joint pdf of $X$ and $Y$, $f(x, y)$ is defined as

$$\frac{d^2}{dxdy} F(x, y) = f(x, y), F(a, b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f(x, y) dxdy$$

$$P\{(x, y) \in D\} = \iint_D f(x, y) dxdy$$

- A variation of Proposition 2.1
  - $E[g(X, Y)] = \sum_x \sum_y g(x, y) p(x, y)$ (discrete case)

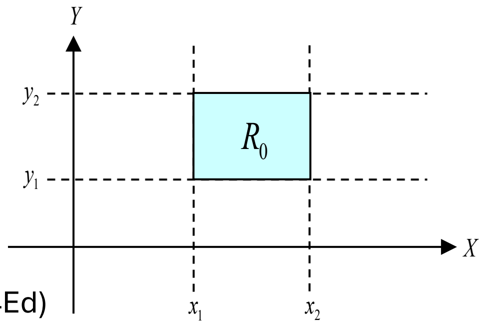
    $= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y)\, dxdy$ (continuous case)
  - E.g) $E[aX + bY] = aE[X] + bE[Y]$

# Properties of Joint Distribution

1. The $F(x, y)$ is such that
$$F(-\infty, y) = 0, F(x, -\infty) = 0, F(\infty, \infty) = 1$$

2. The event $\{x_1 < X \leq x_2, Y \leq y\}$ consists of all points $\{X, Y\}$ in the vertical half-strip $D_2$ and the event $\{\boldsymbol{x} \leq x, y_1 < Y \leq y_2\}$ consists of all points $\{\boldsymbol{x}, \boldsymbol{y}\}$ in the vertical half-strip $D_3$. We maintain that
$$P\{x_1 < X \leq x_2, Y \leq y\} = F(x_2, y) - F(x_1, y)$$
$$P\{X \leq x, y_1 < Y \leq y_2\} = F(x, y_2) - F(x, y_1)$$

*3.* $P\{x_1 < X \leq x_2, Y \leq y\}$
$$= F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$$




From Chap.6 in Papoulis's (4Ed)

# Jointly Distributed RVs: Example 1

- **Example 2.30** As another example of the usefulness of Equation (2.11), let us use it to obtain the expectation of a binomial random variable having parameters $n$ and $p$. Recalling that such a random variable $X$ represents the number of successes in $n$ trials when each trial has probability $p$ of being a success, we have
$$X = X_1 + X_2 + \cdots + X_n$$
where $X_i$ is 1, if the $i$th trial is a success, otherwise 0.
    - $E[X] = np$ from a variation of Proposition 2.1

# Jointly Distributed RVs: Example 2

- **Example 2.31** At a party $N$ men throw their hats into the center of a room. The hats are mixed up and each man randomly selects one. Find the expected number of men who select their own hats.
  - Letting $X$ denote the number of men that select their own hats, we can best compute $E[X]$ by noting that
  $$X = X_1 + X_2 + \cdots + X_N,$$
  where $X_i = 1$, if the $i_{\text{th}}$ man selects his own hat, otherwise 0.
  - Now, because the $i$th man is equally likely to select any of the $N$ hats, it follows that $P\{X_i = 1\} = P\{i_{th}\ man\ selects\ his\ own\ hat\} = 1/N$
  - $E[X] = E[X_i] + E[X_2] + \cdots + E[X_N] = N(1/N) = 1$

# Independent RVs:

- Random variables $X$ and $Y$ are said to be *independent* if, for all $a, b$,
  $P\{X \le a, Y \le b\} = P\{X \le a\}P\{Y \le b\}$
  $\Leftrightarrow F(a,b) = FX(a)FY(b)$ for all $a, b$
  $\Leftrightarrow f(x,y) = f_X(x)f_Y(y)$ for continuous cases
  ($\Leftrightarrow p(x,y) = p_X(x)p_Y(y)$ for discrete cases)

- **Proposition 2.3** If $X$ and $Y$ are independent, then for any functions $h$ and $g$

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

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

# Covariance and Variance of Sums of Random Variables

- The covariance of any two random variables $X$ and $Y$, denoted by $Cov(X,Y)$, is defined by
  - $$\begin{aligned} Cov(X,Y) &= E[(X-E[X])(Y-E[Y])] \\ &= E[XY - YE[X] - XE[Y] + E[X]E[Y]] \\ &= E[XY] - E[Y]E[X] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

- Note that if $X$ and $Y$ are independent, then by Proposition 2.3 it follows that $Cov(X,Y) = 0$.

- **Example 2.33** The joint density function of $X, Y$ is, $f(x,y) = \frac{1}{y}e^{-(y+x/y)}, 0 < x, y < \infty$
  - Verify that the preceding is a joint density function.
  - Find $Cov(X,Y)$.
  - By yourself

# Properties of Covariance

- For any random variables $X, Y, Z$ and constant $c$,
  1. $Cov(X, X) = Var(X)$,
  2. $Cov(X, Y) = Cov(Y, X)$,
  3. $Cov(cX, Y) = c\, Cov(X, Y)$,
  4. $Cov(X, Y + Z) = Cov(X, Y) + Cov(X, Z)$.

  - Proof]
    Whereas the first three properties are immediate, the final one is easily proven as follows:
$$
\begin{aligned}
Cov(X, Y + Z) &= E[X(Y + Z)] - E[X]E[Y + Z] \\
&= E[XY] - E[X]E[Y] + E[XZ] - E[X]E[Z] \\
&= Cov(X, Y) + Cov(X, Z)
\end{aligned}
$$

-

- A useful expression for the variance of the sum of random variables can be obtained as follows:
  - $Var(\sum_{i=1}^{n} X_i) = Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right)$ (2.16)

- If $X_i, i = 1, \ldots, n$ are independent random variables, then Equation (2.16) reduces to
  - $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$

- **Definition 2.1** If $X_1, \ldots, X_n$ are independent and identically distributed, then the random variable $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is called the *sample mean*.

- **Proposition 2.4** Suppose that $X_1, \ldots, X_n$ are independent and identically distributed with expected value $\mu$ and variance $\sigma^2$. Then,
  a. $E[\bar{X}] = \mu$
  b. $\mathrm{Var}(\bar{X}) = \sigma^2/n$
  c. $\mathrm{Cov}(\bar{X}, X_i - \bar{X}) = 0$, i=1, ...,n
  - Proofs...

# One Function of Two RVs

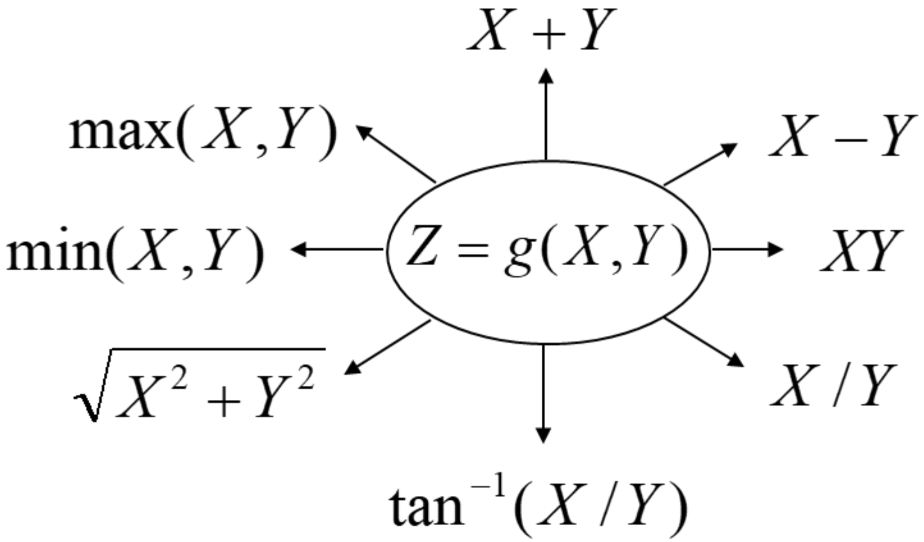
- Given two random variables $X$ and $Y$ and a function $g(X, Y)$, we form a new random variable $Z$ as $Z = g(X, Y)$.

- Given the joint p.d.f $f_{XY}(X, Y)$, how does one obtain $f_Z(Z)$ the p.d.f of $Z$?
  - $f_Z(Z) = P\{Z(\zeta) \leq z\} = P\{g(X, Y) \leq z\} = P\{(X, Y) \in D_z\} = \iint_{D_Z} f(x, y)dxdy$

- Problems of this type are of interest from a practical standpoint. For example, a receiver output signal usually consists of the desired signal buried in noise, and the above formulation in that case reduces to $Z = X + Y$.

From Chap.6 in Papoulis's (4Ed)

# Examples of One Function of Two RVs




From Chap.6 in Papoulis's (4Ed)

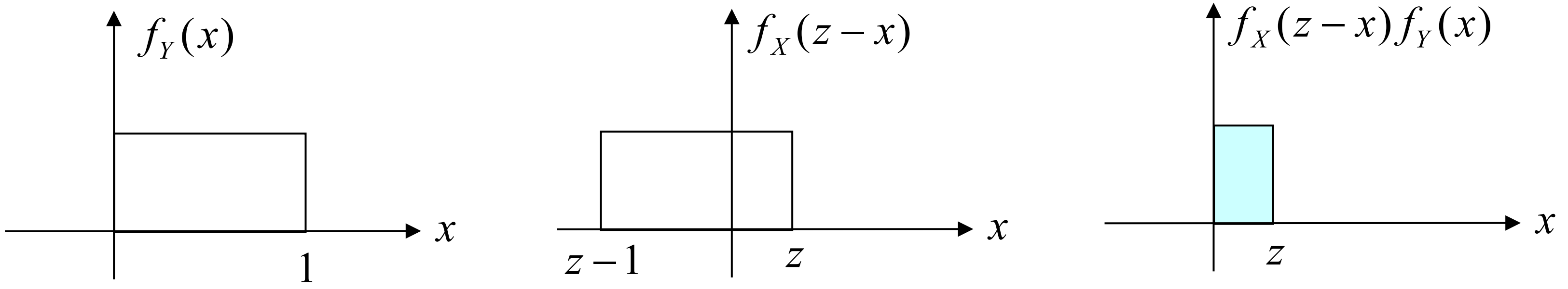
# Distribution of X+Y

- Suppose first that $X$ and $Y$ are continuous, $X$ having probability density $f$ and $Y$ having probability density $g$. Then, letting $F_{X+Y}(a)$ be the cumulative distribution function of $X+Y$, we have
  - $F_{X+Y}(a) = P\{X+Y \le a\} = \int\int_{x+y\le a} f(x)g(y)\,dx\,dy$
    $= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f(x)g(y)\,dx\,dy \;= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{a-y} f(x)\,dx\right)g(y)\,dy$
    $= \int_{-\infty}^{\infty} F_X(a-y)g(y)\,dy$
    - The cumulative distribution function $F_{X+Y}$ is called the *convolution* of the distributions $F_X$ and $F_Y$.

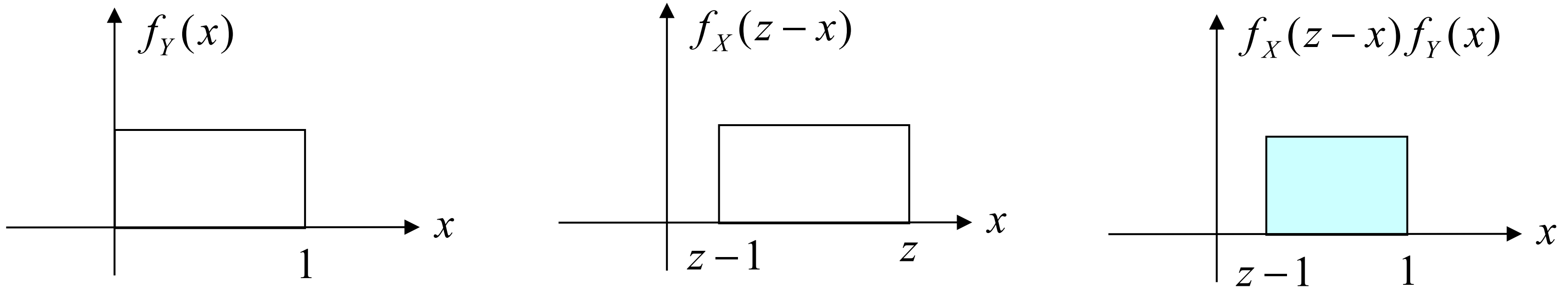
  - $f_{X+Y}(a) = \frac{d}{da}\int_{-\infty}^{\infty} F_X(a-y)g(y)\,dy$
    $= \int_{-\infty}^{\infty} \frac{d}{da} F_X(a-y)g(y)\,dy = \int_{-\infty}^{\infty} f(a-y)g(y)\,dy$

# One Function of Two RVs: Example 1

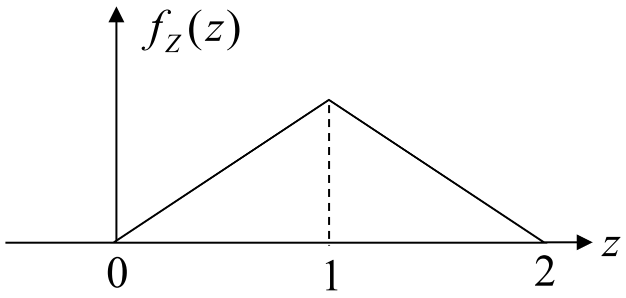
- **Example 2.36 (Sum of Two Independent Uniform Random Variables)** If $X$ and $Y$ are independent random variables both uniformly distributed on (0, 1), then calculate the probability density of $X + Y$.
  - Answer?




$(a)\quad 0 \le z < 1$



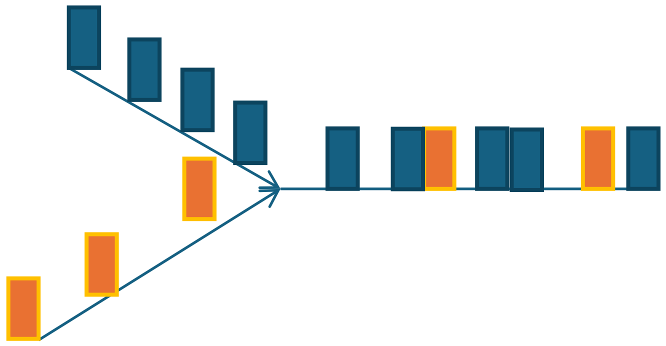

$(b)\quad 1 \le z < 2$




From Chap.6 in Papoulis's (4Ed)

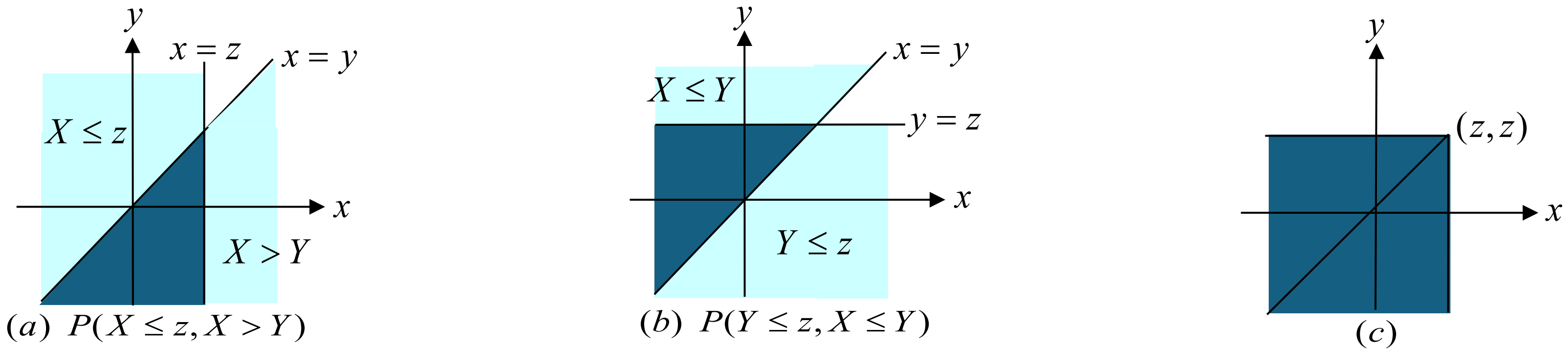
# One Function of Two RVs: Example 2

- **Example 2.37 (Sums of Independent Poisson Random Variables)** Let $X$ and $Y$ be independent Poisson random variables with respective means $\lambda_1$ and $\lambda_2$.
  Calculate the distribution of $X + Y$.
    - Homework



# One Function of Two RVs: Example 3

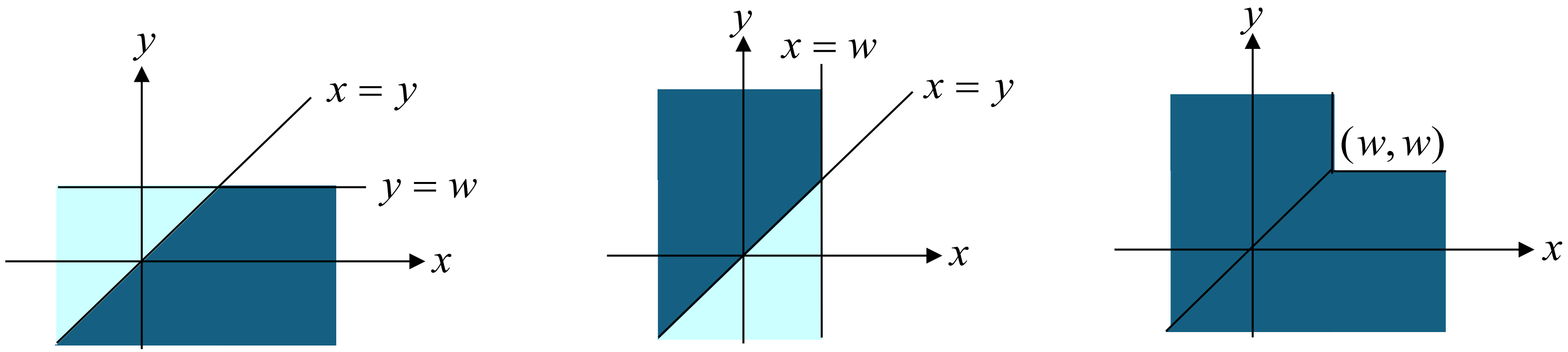
- Let $Z = max(X, Y)$. Determine $f_Z(z)$
  - Homework




From Chap.6 in Papoulis's (4Ed)

# One Function of Two RVs: Example 4

- Let $W = min(X, Y)$. Determine $f_W(w)$
  - Homework



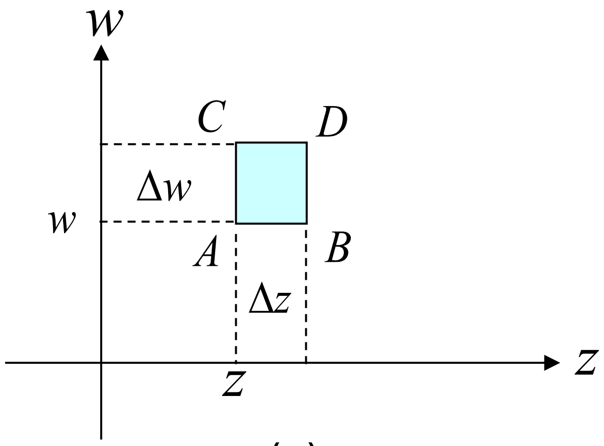

From Chap.6 in Papoulis's (4Ed)

# Joint Probability Distribution of Functions of Random Variables

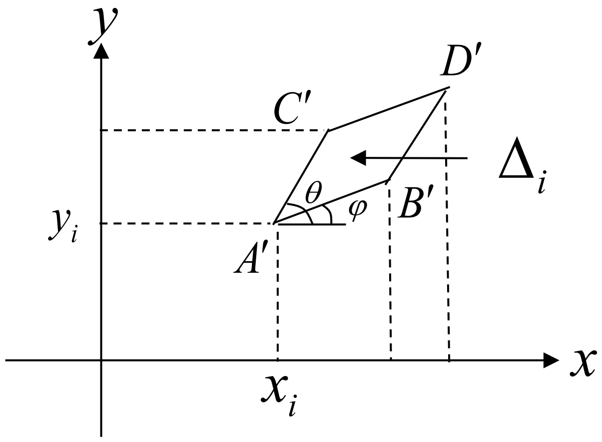
- Let $X_1$ and $X_2$ be jointly continuous random variables with joint probability density function $f(x_1, x_2)$.
- Suppose $Y_1 = g_1(x_1, x_2), Y_2 = g_2(x_1, x_2)$.
  - $F_{Y_1,Y_2}(y_1, y_2) = \iint_{D_{y_1,y_2}} f_{Y_1,Y_2}(y_1, y_2) dy_1 dy_2$
    $= P\{Y_1 \le y_1, Y_2 \le y_2\} = P\{g_1(x_1, x_2) \le y_1, g_1(x_1, x_2) \le y_2\}$
    $= P\{(x_1, x_2) \in D_{Y_1,Y_2}\} = \iint_{D_{Y_1,Y_2}} f_{X_1,X_2}(x_1, x_2) dx_1 dx_2$
    $= \iint_{D_{X_1,X_2}} f_{X_1,X_2}(x_1, x_2) |J(x_1, x_2)|^{-1} dx_1 dx_2$
  - $f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(x_1, x_2) |J(x_1, x_2)|^{-1}$

    where $|J(x_1, x_2)| = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix} = |J(y_1, y_2)|^{-1}$

# Meaning of $|J(x_1, x_2)|$




(a)




(b)

# Example of TWO RVs

- Suppose $X$ and $Y$ are zero mean independent Gaussian random variables with common variance $\sigma^2$. Define $R = \sqrt{X^2 + Y^2}$, $\Theta = \tan^{-1}\left(\frac{Y}{X}\right)$, where $|\Theta| < \pi$. Find the joint density function
  - $f_{X,Y}(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}, r = \sqrt{x^2 + y^2}, \theta = \tan^{-1}\left(\frac{y}{x}\right), x = r\cos(\theta), y = r\sin(\theta)$
  - $|J(r,\theta)| = \left\| \begin{vmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \end{vmatrix} \right\| = \left\| \begin{vmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{vmatrix} \right\| = r = |J(x,y)|^{-1}$
  - $f_{R,\Theta}(r,\theta) = f_{X,Y}(x,y)|J(x,y)|^{-1} = r f_{X,Y}(x,y) = \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}}$, for $0 < r < \infty$, $|\theta| < \pi$
  - $f_R(r) = \int_{-\pi}^{\pi} f_{R,\Theta}(r,\theta)d\theta = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}, 0 < r < \infty$
    $f_\Theta(\theta) = \int_0^{\infty} f_{R,\Theta}(r,\theta)dr = \frac{1}{2\pi}, |\theta| < \pi$
  - $f_{R,\Theta}(r,\theta) = f_R(r) f_\Theta(\theta)$

From Chap.6 in Papoulis's (4Ed)

# Moment Generating Functions (1/2)

- The *moment generating function $\phi(t)$* of the random variable *X* is defined for all values *t* by
  - $\phi(t) = E[e^{tX}] = \begin{cases} \sum_x e^{tx} p(x) \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx \end{cases}$

- Properties
  - $\phi'(t) = \frac{d}{dt} E[e^{tX}] = E\left[\frac{d}{dt} e^{tX}\right] = E[Xe^{tX}]$
  - $\phi''(t) = \frac{d}{dt} \phi'(t) = E\left[\frac{d}{dt} (Xe^{tX})\right] = E[X^2 e^{tX}]$
  - $\phi^{(n)}(t) = E[X^n e^{tX}]$
    - $\phi'(0) = E[X], \phi''(0) = E[X^2], \phi^{(n)}(0) = E[X^n]$

# Moment Generating Functions (2/2)

- An important property of moment generating functions is that the *moment generating function of the sum of independent random variables is just the product of the individual moment generating functions*. To see this, suppose that $X$ and $Y$ are independent and have moment generating functions $\phi_X(t)$ and $\phi_Y(t)$, respectively. Then $\phi_{X+Y}(t)$, the moment generating function of $X+Y$, is given by
    - $\phi_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX}e^{tY}] = E[e^{tX}]E[e^{tY}] = \phi_X(t)\phi_Y(t)$

| Discrete probability distribution | Probability mass function, $p(x)$ | Moment generating function, $\phi(t)$ | Mean | Variance |
|---|---|---|---|---|
| Binomial with parameters $n, p$, $0 \le p \le 1$ | $\binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n$ | $(pe^t + (1-p))^n$ | $np$ | $np(1-p)$ |
| Poisson with parameter $\lambda > 0$ | $e^{-\lambda} \frac{\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$ | $\exp\{\lambda(e^t - 1)\}$ | $\lambda$ | $\lambda$ |
| Geometric with parameter $0 \le p \le 1$ | $p(1-p)^{x-1}$, $x = 1, 2, \ldots$ | $\frac{pe^t}{1-(1-p)e^t}$ | $\frac{1}{p}$ | $\frac{1-p}{p^2}$ |

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

# Examples of MGFs

- Computation of the mean and variance of a RV using its moment generating function
  - Example 2.40 (The Binomial Distribution with Parameters n and p)
  - Example 2.41 (The Poisson Distribution with Mean λ)
  - Example 2.42 (The Exponential Distribution with Parameter λ)
  - Example 2.43 (The Normal Distribution with Parameters μ and σ2)

- Sums of Independent Random Variables
  - Example 2.44 (Sums of Independent Binomial Random Variables)
  - Example 2.45 (Sums of Independent Poisson Random Variables)
  - Example 2.46 (Sums of Independent Normal Random Variables)

# Limit Theorems: Markov's Inequality

- Proposition 2.6 (Markov's Inequality) If $X$ is a random variable that takes only nonnegative values, then for any value $a > 0$
$$P\{X \geq a\} \leq (E[X])/a$$
- Proof

-

# Limit Theorems: Chebyshev's Inequality

- Proposition 2.7 (Chebyshev's Inequality) If X is a random variable with mean $\mu$ and variance $\sigma^2$, then, for any value $k > 0$,

$$P\{|X-\mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

- Proof
- Hint: Markov's inequality and |X-u|2 ≥k2

-

# Example of Inequalities

- Example 2.49 Suppose we know that the number of items produced in a factory during a week is a random variable with mean 500.
  - What can be said about the probability that this week's production will be at least 1000?
  - If the variance of a week's production is known to equal 100, then what can be said about the probability that this week's production will be between 400 and 600?

# Limit Theorems: Strong Law of Large Numbers

- **Theorem 2.1 (Strong Law of Large Numbers)** Let $X_1, X_2, ...$ be a sequence of independent random variables having a common distribution, and let $E[Xi] = \mu$. Then, with probability 1, $\frac{X_1+X_2+\cdots+X_n}{n} \to \mu$

- **Example:** suppose that a sequence of independent trials is performed. Let $E$ be a fixed event and denote by $P(E)$ the probability that $E$ occurs on any particular trial. Letting
$$X_i = \begin{cases} 1, if\ E\ occurs\ on\ the\ ith\ trial \\ 0, if\ E\ does\ not\ occur\ on\ the\ ith\ trial \end{cases}$$
we have by the strong law of large numbers that, with probability 1,
$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to E[X] = P(E)$$

# Limit Theorems: Central Limit Theorem

- **Theorem 2.2 (Central Limit Theorem)** Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed random variables, each with mean μ and variance $\sigma^2$. Then the distribution of $\frac{X_1+X_2+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$ tends to the standard normal as n→∞. That is $\mathrm{P}\left\{\frac{X_1+X_2+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \le a\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-\frac{x^2}{2}}dx$ as n→∞.
- If $X$ is binomially distributed with parameters $n$ and $p$, then $X$ has the same distribution as the sum of $n$ independent Bernoulli random variables, each with parameters $p$. Hence, the distribution of

$$\frac{X-E[X]}{\sqrt{Var(X)}} = \frac{X-np}{\sqrt{np(1-p)}}$$

approaches the standard normal distribution as $n$ approaches ∞.

# Stochastic Processes

- A stochastic process $\{X(t), t \in T\}$ is a collection of random variables. That is, for each $t \in T$, $X(t)$ is a random variable. The index t is often interpreted as time and, as a result, we refer to $X(t)$ as the state of the process at time $t$.
  - (e.g) $X(t)$ : the total number of customers that have entered a supermarket by time $t$
- The set $T$ is called the index set of the process.
  - $T$ is a countable set: a discrete-time process.
  - $T$ is an interval of the real line: a continuous-time process.
- The state space of a stochastic process is defined as the set of all possible values that the random variables $X(t)$ can assume. Thus, a stochastic process is a family of random variables that describes the evolution through time of some (physical) process. We shall see much of stochastic processes in the following chapters of this text.