# Random Variables <br> Stochastic Processes - Lecture Notes 

## Fatih Cavdur

to accompany
Introduction to Probability Models by Sheldon M. Ross

Fall 2015

## Outline I

Random Variables

Discrete RVs

Continuous RVs

Expectation of an RV

Jointly Distributed RVs

Moment Generating Function

Limit Theorems

## Outline II

Stochastic Processes

## Random Variables

Sometimes, in performing an experiment, we are interested in some functions of the outcome as opposed to the outcome itself. These quantities of interest, or more formally, these real-valued functions defined on the sample space, are known as random variables (RVs).

## Example

We let $X$ denote the RV that is defined as the sum of two fair dice. We then have

$$
\begin{aligned}
P\{X=2\} & =P\{(1,1)\}=\frac{1}{36} \\
P\{X=3\} & =P\{(1,2),(2,1)\}=\frac{2}{36} \\
P\{X=4\} & =P\{(1,3),(2,2)(3,1)\}=\frac{3}{36} \\
\cdots & =\cdots \\
P\{X=12\} & =P\{(6,6)\}=\frac{1}{36}
\end{aligned}
$$

## Example (cont.)

We also note that

$$
P\left\{\bigcup_{i=2}^{12} X=i\right\}=\sum_{i=2}^{12} P\{X=i\}=1
$$

## Example

We toss 2 coins. Let $Y$ be the number of heads appearing. We then have

$$
\begin{aligned}
& P\{Y=0\}=P\{(t, t)\}=\frac{1}{4} \\
& P\{Y=1\}=P\{(h, t),(t, h)\}=\frac{2}{4} \\
& P\{Y=2\}=P\{(h, h)\}=\frac{1}{4}
\end{aligned}
$$

We also note that $P\{Y=0\}+P\{Y=1\}+P\{Y=2\}=1$.

## Example

Suppose that we toss a coin having a probability $p$ of coming heads until the first head appears. Let $N$ be the number of flips required, and assuming that the outcomes of successive flips are independent, $N$ is an RV with

$$
\begin{aligned}
& P\{N=1\}=P\{h\}=p \\
& P\{N=2\}=P\{(t, h)\}=(1-p) p \\
& P\{N=3\}=P\{(t, t, h)\}=(1-p)^{2} p
\end{aligned}
$$

## Example (cont.)

In general,

$$
P\{N=n\}=P\{t, t, \ldots, t, h\}=(1-p)^{n-1} p, \quad n \geq 1
$$

We also note that

$$
\begin{aligned}
P\left(\bigcup_{n=1}^{\infty} N=n\right) & =\sum_{n=1}^{\infty} P\{N=n\} \\
& =p \sum_{n=1}^{\infty}(1-p)^{n-1} \\
& =\frac{p}{1-(1-p)} \\
& =1
\end{aligned}
$$

## Discrete and Continuous RVs

- RVs that take on either a finite or a countable number of values are called discrete RVs, such as the number of heads or tails when we flip 2 coins as in the above example
- RVs that take on continuous values are called continuous RVs, such as the lifetime of a car.


## Cumulative Distribution Function

The Cumulative Distribution Function (CDF) $F$ of an RV $X$ is defined for any real number $b$ as the probability of $X$ takes on a value that is less than or equal to $b$.

$$
F(b)=P\{X \leq b\}
$$

- $F(b)$ is a non-decreasing function of $b$.
- $\lim _{b \rightarrow-\infty} F(b)=F(-\infty)=0$
- $\lim _{b \rightarrow+\infty} F(b)=F(+\infty)=1$

We can answer all probability questions about $X$ in terms of the CDF F. How?

## Discrete RVs

We define the Probability Mass Function (PMF) of $X$ as

$$
p(a)=P\{X=a\}
$$

We also have

$$
\sum_{i=1}^{\infty} p\left(x_{i}\right)=1
$$

## Bernoulli RV

- X: 1 (success) or 0 (failure)
- $p$ : probability of success

The PMF of $X$ is

$$
p(i)=P\{X=i\}=p^{i}(1-p)^{1-i}
$$

where $i=0,1$.

## Binomial RV

- $X$ : number of successes
- $n$ : number of trials
- p: probability of success in each trial

The PMF of $X$ is

$$
p(i)=\binom{n}{i} p^{i}(1-p)^{n-i}, \quad i=0,1,2, \ldots, n
$$

We note that

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

## Example

It is known that a 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 3 items, at most 1 will be defective?

Let $X$ be the number of defective items.

$$
P\{X=0\}+P\{X=1\}=\binom{3}{0}(.1)^{0}(.9)^{3}+\binom{3}{1}(.1)^{1}(.9)^{2}=.972
$$

## Example

Suppose that an airplane engine will fail, when in flight, with probability $1-p$, independently from the engine to engine. If the airplane makes a successful flight when at least $50 \%$ of its engines are operative, for what $p$ values a 4-engine airplane is safer than a 2-engine airplane?

## Example (cont.)

Suppose that an airplane engine will fail, when in flight, with probability $1-p$, independently from the engine to engine. If the airplane makes a successful flight when at least $50 \%$ of its engines are operative, for what $p$ values a 4-engine airplane is safer than a 2-engine airplane?

Let $X$ and $Y$ be the number of operative engines for 4 and 2-engine airplanes, respectively.

## Example (cont.)

For a 4-engine airplane, we have

$$
\begin{aligned}
\sum_{i=2}^{4} P\{X=i\} & =\binom{4}{2} p^{2}(1-p)^{2}+\binom{4}{3} p^{3}(1-p)^{1}+\binom{4}{4} p^{4}(1-p)^{0} \\
& =p^{4}+4 p^{3}(1-p)+6 p^{2}(1-p)^{2}
\end{aligned}
$$

For a 2-engine airplane, we have

$$
\begin{aligned}
\sum_{i=1}^{2} P\{X=i\} & =\binom{2}{1} p^{1}(1-p)^{1}+\binom{2}{2} p^{2}(1-p)^{0} \\
& =p^{2}+2 p(1-p)
\end{aligned}
$$

## Example (cont.)

A 4-engine airplane is safer if

$$
\begin{aligned}
p^{4}+4 p^{3}(1-p)+6 p^{2}(1-p)^{2} & \geq p^{2}+2 p(1-p) \\
p^{3}+4 p^{2}(1-p)+6 p(1-p)^{2} & \geq-p+2 \\
3 p^{3}-8 p^{2}+7 p-2 & \geq 0 \\
(p-1)^{2}(3 p-2) & \geq 0 \\
(3 p-2) & \geq 0 \\
p & \geq \frac{2}{3}
\end{aligned}
$$

## Geometric RV

- X: number of trials until the first success
- $p$ : probability of success in each trial

The PMF of $X$ is

$$
p(i)=(1-p)^{i-1} p, \quad i=1,2, \ldots
$$

We also have

$$
\sum_{i=1}^{\infty} p(i)=p \sum_{i=1}^{\infty}(1-p)^{n-1}=\frac{p}{1-(1-p)}=1
$$

## Poisson RV

- X: number of outcomes during a unit time period
- $\lambda$ : rate

The PMF of $X$, for $\lambda>0$ is

$$
p(i)=\frac{e^{-\lambda} \lambda^{i}}{i!}, \quad i=0,1, \ldots
$$

We also note that

$$
\sum_{i=0}^{\infty} p(i)=e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^{i}}{i!}=e^{-\lambda} e^{\lambda}=1
$$

## Poisson Approximation to Binomial

When $n$ is large and $p$ is small, a binomial RV can be approximated by Poisson.

Let $X$ be Binomial with $(n, p)$ and $\lambda=n p$.

$$
\begin{aligned}
p(i) & =\frac{n!}{(n-i)!i!} p^{i}(1-p)^{n-i} \\
& =\frac{n!}{(n-i)!i!}\left(\frac{\lambda}{n}\right)^{i}\left(1-\frac{\lambda}{n}\right)^{n-i} \\
& =\frac{n(n-1) \ldots(n-i+1)}{n^{i}} \frac{\lambda^{i}}{i!} \frac{\left(1-\frac{\lambda}{n}\right)^{n}}{\left(1-\frac{\lambda}{n}\right)^{i}}
\end{aligned}
$$

## Poisson Approximation to Binomial

Let $X$ be Binomial with $(n, p)$ and $\lambda=n p$.

$$
\begin{aligned}
p(i) & =\frac{n!}{(n-i)!i!} p^{i}(1-p)^{n-i} \\
& =\frac{n!}{(n-i)!i!}\left(\frac{\lambda}{n}\right)^{i}\left(1-\frac{\lambda}{n}\right)^{n-i} \\
& =\frac{n(n-1) \ldots(n-i+1)}{n^{i}} \frac{\lambda^{i}}{i!} \frac{\left(1-\frac{\lambda}{n}\right)^{n}}{\left(1-\frac{\lambda}{n}\right)^{i}}
\end{aligned}
$$

Since $n$ is large and $p$ is small, we have

$$
\left(1-\frac{\lambda}{n}\right)^{n} \approx e^{-\lambda}, \quad \frac{n(n-1) \ldots(n-i+1)}{n^{i}} \approx 1, \quad\left(1-\frac{\lambda}{n}\right)^{i} \approx 1
$$

## Poisson Approximation to Binomial

As a result,

$$
p(i)=\frac{n(n-1) \ldots(n-i+1)}{n^{i}} \frac{\lambda^{i}}{i!} \frac{\left(1-\frac{\lambda}{n}\right)^{n}}{\left(1-\frac{\lambda}{n}\right)^{i}}
$$

Since $n$ is large and $p$ is small, we have

$$
\left(1-\frac{\lambda}{n}\right)^{n} \approx e^{-\lambda}, \quad \frac{n(n-1) \ldots(n-i+1)}{n^{i}} \approx 1, \quad\left(1-\frac{\lambda}{n}\right)^{i} \approx 1
$$

and thus, we have

$$
p(i) \approx e^{-\lambda} \frac{\lambda^{i}}{i!}
$$

## Continuous RVs

$X$ is 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

$$
P\{X \in B\}=\int_{B} f(x) d x
$$

The function $f(x)$ is called the probability density function (PDF) of the RV $X$. The probability that $X$ is in $B$ can be computed by integrating the PDF over set $B$.

## Continuous RVs

We have

$$
P\{X \in(-\infty,+\infty)\}=\int_{-\infty}^{+\infty} f(x) d x=1
$$

and the probability that $X \in[a, b]$ is computed as

$$
P\{a \leq X \leq b\}=\int_{a}^{b} f(x) d x
$$

What is the difference between $P\{a \leq X \leq b\}$ and $P\{a<X<b\}$ ?

## Continuous RVs

The cumulative distribution function of $X$ is

$$
F(a)=P\{X \in(-\infty, a]\}=\int_{-\infty}^{a} f(x) d x
$$

Differentiating the preceding, we obtain

$$
\frac{d}{d a} F(a)=f(a)
$$

## Uniform RV

Let $X$ be a uniform RV on the interval ( $\alpha, \beta$ ). Its PDF

$$
f(x)=\left\{\begin{array}{cc}
\frac{1}{\beta-\alpha}, & \alpha<x<\beta \\
0, & \text { otherwise }
\end{array}\right.
$$

The CDF of a uniform RV $X$ is

$$
F(x)=\left\{\begin{array}{cc}
0, & x \leq \alpha \\
\frac{x-\alpha}{\beta-\alpha}, & \alpha<x<\beta \\
1, & x \geq \beta
\end{array}\right.
$$

## Exponential RV

Let $X$ be an exponential RV with parameter $\lambda$. Its PDF

$$
f(x)=\left\{\begin{array}{cc}
\lambda e^{-\lambda x}, & x \geq 0 \\
0, & \text { otherwise }
\end{array}\right.
$$

The CDF of an exponential RV $X$ is

$$
F(x)=\left\{\begin{array}{cc}
1-e^{-\lambda x}, & x \geq 0 \\
0, & \text { otherwise }
\end{array}\right.
$$

## Gamma RV

Let $X$ be a gamma RV with parameters $\lambda$ and $\alpha$. Its PDF

$$
f(x)=\left\{\begin{array}{cc}
\frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & x \geq 0 \\
0, & \text { otherwise }
\end{array}\right.
$$

for some $\alpha>0$ and $\lambda>0$. The gamma function is defined as

$$
\Gamma(\alpha)=\int_{0}^{\infty} e^{-x} x^{\alpha-1} d x
$$

and we can show that, for some integer $n$,

$$
\Gamma(n)=(n-1)!
$$

## Normal RV

Let $X$ be a normal RV with parameters $\mu$ and $\sigma^{2}$. Its PDF

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

for $x \in(-\infty,+\infty)$.

## Normal RV

If $X$ is normally distributed with $\mu$ and $\sigma^{2}$, then, $Y=\alpha X+\beta$ is normally distributed with parameters $\alpha \mu+\beta$ and $\alpha^{2} \sigma^{2}$. How to prove that?

## Normal RV

Suppose that $\alpha>0$, and the CDF of $Y$ is as follows:

$$
\begin{aligned}
F_{Y}(a) & =P\{Y \leq a\} \\
& =P\{\alpha X+\beta \leq a\} \\
& =P\left\{X \leq \frac{a-\beta}{\alpha}\right\} \\
& =F_{X}\left(\frac{a-\beta}{\alpha}\right) \\
& =\int_{-\infty}^{(a-\beta) / \alpha} \frac{1}{\sqrt{2 \pi \sigma^{2}}} e^{-\frac{(x-\mu)^{2}}{2 \sigma^{2}}} d X \quad(\text { set } v=\alpha x+\beta) \\
& =\int_{-\infty}^{a} \frac{1}{\sqrt{2 \pi \sigma^{2} \alpha^{2}}} e^{-\frac{(v-(\alpha \mu+\beta))^{2}}{2 \sigma^{2} \alpha^{2}}} d v
\end{aligned}
$$

## Normal RV

Since we have

$$
F_{Y}(a)=\int_{-\infty}^{a} \frac{1}{\sqrt{2 \pi \sigma^{2} \alpha^{2}}} e^{-\frac{[v-(\alpha \mu+\beta))^{2}}{2 \sigma^{2} \alpha^{2}}} d v
$$

we can write,

$$
f_{Y}(v)=\frac{1}{\sqrt{2 \pi \sigma^{2} \alpha^{2}}} e^{-\frac{[v-(\alpha \mu+\beta)]^{2}}{2 \sigma^{2} \alpha^{2}}}
$$

Hence, $Y$ is normally distributed with parameters $\alpha \mu+\beta$ and $\alpha^{2} \sigma^{2}$.

$$
X \sim \mathrm{~N}\left(\mu, \sigma^{2}\right) \Rightarrow Y=\alpha X+\beta \sim \mathrm{N}\left(\alpha \mu+\beta, \alpha^{2} \sigma^{2}\right)
$$

## Standard Normal RV

We proved that if $X$ is normally distributed with $\mu$ and $\sigma^{2}$, then, $Y=\alpha X+\beta$ is normally distributed with parameters $\alpha \mu+\beta$ and $\alpha^{2} \sigma^{2}$.

$$
X \sim \mathrm{~N}\left(\mu, \sigma^{2}\right) \Rightarrow Y=\alpha X+\beta \sim \mathrm{N}\left(\alpha \mu+\beta, \alpha^{2} \sigma^{2}\right)
$$

An implication of the preceding result is that if $X$ is normally distributed with parameters $\mu$ and $\sigma^{2}$, then,

$$
Z=\frac{X-\mu}{\sigma}
$$

is normally distributed with parameters 0 and 1 , and such an RV $Y$ is said to be a standard normal RV.

## Expectation of a Discrete RV

Expectation of a discrete RV $X$ is defined as

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

If we let $X$ be the outcome when we roll a die,

$$
E(X)=\sum_{x=1}^{6} x p(x)=\frac{1}{6} \sum_{x=1}^{6} x=\frac{(6)(7)}{(6)(2)}=\frac{7}{2}
$$

Expectation of a Binomial RV

$$
\begin{aligned}
E(X) & =\sum_{i=0}^{n}\binom{n}{i} p^{i}(1-p)^{n-i} \\
& =\sum_{i=1}^{n} \frac{i n!}{(n-i)!!!} p^{i}(1-p)^{n-i} \\
& =\sum_{i=1}^{n} \frac{n!}{(n-i)!(i-1)!} p^{i}(1-p)^{n-i} \\
& =n p \sum_{i=1}^{n} \frac{(n-1)!}{(n-i)!(i-1)!} p^{i-1}(1-p)^{n-i} ; \text { let } k=i-1 \\
& =n p \sum_{k=0}^{n-1}\binom{n-1}{k} p^{k}(1-p)^{n-1-k} \\
& =n p[p+(1-p)]^{n-1}=n p
\end{aligned}
$$

Expectation of a Geometric RV

$$
\begin{aligned}
E(X) & =\sum_{i=1}^{\infty} i p(1-p)^{i-1} ; \text { let } q=1-p \\
& =p \sum_{i=1}^{\infty} i q^{i-1} \\
& =p \sum_{i=1}^{\infty} \frac{d q^{i}}{d q} \\
& =p \frac{d}{d q}\left(\sum_{i=1}^{\infty} q^{n}\right) \\
& =p \frac{d}{d q}\left(\frac{q}{1-q}\right) \\
& =\frac{p}{(1-q)^{2}}=\frac{1}{p}
\end{aligned}
$$

Expectation of a Poisson RV

$$
\begin{aligned}
E(X) & =\sum_{i=0}^{\infty} \frac{i e^{-\lambda} \lambda^{i}}{i!} \\
& =\sum_{i=1}^{\infty} \frac{e^{-\lambda} \lambda^{i}}{(i-1)!} \\
& =\lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\
& =\lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k}}{k!} \\
& =\lambda e^{-\lambda} e^{\lambda} \\
& =\lambda
\end{aligned}
$$

Expectation of a Continuous RV
Expectation of a discrete RV $X$ is defined as

$$
E(X)=\int_{-\infty}^{+\infty} x f(x) d x
$$

Expectation of a Uniform RV

$$
\begin{aligned}
E(X) & =\int_{-\infty}^{+\infty} x f(x) d x \\
& =\int_{\alpha}^{\beta} \frac{x d x}{\beta-\alpha} \\
& =\frac{\beta^{2}-\alpha^{2}}{\beta-\alpha} \\
& =\frac{\alpha+\beta}{2}
\end{aligned}
$$

Expectation of an Exponential RV

$$
\begin{aligned}
E(X) & =\int_{-\infty}^{+\infty} x f(x) d x \\
& =\int_{0}^{\infty} x \lambda e^{-\lambda x} d x \\
& =\left.\left(-x e^{-\lambda x}\right)\right|_{0} ^{\infty}-\int_{0}^{\infty} e^{-\lambda x} d x \\
& =0-\left.\left(\frac{e^{-\lambda x}}{\lambda}\right)\right|_{0} ^{\infty} \\
& =\frac{1}{\lambda}
\end{aligned}
$$

## Expectation of a Normal RV

$$
\begin{aligned}
E(X) & =\int_{-\infty}^{+\infty} x f(x) d x \\
& =\frac{1}{\sqrt{2 \pi \sigma^{2}}} \int_{-\infty}^{+\infty} x e^{-\frac{(x-\mu)^{2}}{2 \sigma^{2}}} d x ; \text { let } x=(x-\mu)+\mu \\
& =\frac{1}{\sqrt{2 \pi \sigma^{2}}} \int_{-\infty}^{+\infty}(x-\mu) e^{-\frac{(x-\mu)^{2}}{2 \sigma^{2}}} d x \\
& +\frac{\mu}{\sqrt{2 \pi \sigma^{2}}} \int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^{2}}{2 \sigma^{2}}} d x ; \text { let } y=(x-\mu) \\
& =\frac{1}{\sqrt{2 \pi \sigma^{2}}} \int_{-\infty}^{+\infty} y e^{-\frac{y^{2}}{2 \sigma^{2}}} d y+\mu \int_{-\infty}^{+\infty} f(x) d x \\
& =0+\mu \\
& =\mu
\end{aligned}
$$

## Expectation of a Function of an RV

Proposition
Proposition
If $X$ is discrete RV with a PMF $p(x)$, then, for any real-valued function $g$, we have

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

If $X$ is continuous RV with a PDF $f(x)$, then, for any real-valued function $g$, we have

$$
E[g(x)]=\int_{-\infty}^{+\infty} g(x) f(x) d x
$$

## Expectation of a Function of an RV

Corollary
If $a$ and $b$ are constants, then, we have

$$
E(a X+b)=a E(x)+b
$$

Proof (for the Discrete Case)

$$
\begin{aligned}
E(a X+b) & =\sum_{x: p(x)>0}(a x+b) p(x) \\
& =a \sum_{x: p(x)>0} x p(x)+\sum_{x: p(x)>0} p(x) \\
& =a E(X)+b
\end{aligned}
$$

## Moment of an RV

- The expected value of an RV is also referred as the mean or the first moment of the RV.
- $E(X)$ is called as the first moment of $X$.
- In general, $E\left(X^{n}\right), n \geq 1$ is called as the $n$th moment of $X$.


## Moment of an RV

If $X$ is discrete,

$$
E\left(X^{n}\right)=\sum_{x: p(x)>0} x^{n} p(x)
$$

If $X$ is continuous,

$$
E\left(X^{n}\right)=\int_{-\infty}^{+\infty} x^{n} f(x) d x
$$

## Variance

Another quantity of interest is the variance of an RV defined as follows:

$$
\operatorname{Var}(X)=E\left\{[X-E(X)]^{2}\right\}
$$

It measures the expected square of the deviation of the RV from its expected value.

## Another Variance Expression

$$
\begin{aligned}
\operatorname{Var}(X) & =E\left[(X-\mu)^{2}\right] \\
& =E\left(X^{2}-2 \mu X+\mu^{2}\right) \\
& =\int_{-\infty}^{+\infty}\left(x^{2}-2 \mu x+\mu^{2}\right) f(x) d x \\
& =\int_{-\infty}^{+\infty} x^{2} f(x) d x-2 \mu \int_{-\infty}^{+\infty} x f(x) d x+\mu^{2} \int_{-\infty}^{+\infty} f(x) d x \\
& =E\left(X^{2}\right)-2 \mu^{2}+\mu^{2} \\
& =E\left(X^{2}\right)-\mu^{2}
\end{aligned}
$$

Prove for the discrete case!

## Jointly Distributed RVs

We define, for any two RVs $X$ and $Y$, the joint cumulative distribution function of $X$ and $Y$ as

$$
F(a, b)=P\{X \leq a, Y \leq b\}, \quad-\infty<a, b<+\infty
$$

The distribution of $X$ and $Y$ alone can be obtained from the joint distribution as follows:

$$
\begin{aligned}
& F_{X}(a)=P\{X \leq a\}=P\{X \leq a, Y<\infty\}=F(a, \infty) \\
& F_{Y}(b)=P\{Y \leq b\}=P\{X<\infty, Y \leq b\}=F(\infty, b)
\end{aligned}
$$

## Jointly Distributed RVs

## For two discrete $\mathrm{RVs} X$ and $Y$, we can write

$$
p(x, y)=P\{X=x, Y=y\}
$$

The PMF of $X$ and $Y$ are

$$
\begin{aligned}
& p_{X}(x)=\sum_{y: p(x, y)>0} p(x, y) \\
& p_{Y}(y)=\sum_{x: p(x, y)>0} p(x, y)
\end{aligned}
$$

## Jointly Distributed RVs

For two continuous $\mathrm{RVs} X$ and $Y$, we can write

$$
P\{X \in A, Y \in B\}=\int_{B} \int_{A} f(x, y) d x d y
$$

The PMF of $X$ and $Y$ are

$$
\begin{aligned}
P\{X \in A\}=P\{X \in A, Y \in(-\infty,+\infty)\} & =\int_{-\infty}^{+\infty} \int_{A} f(x, y) d x d y \\
& =\int_{A} f_{X}(x) d x \\
P\{Y \in B\}=P\{X \in(-\infty,+\infty), Y \in B\} & =\int_{B} \int_{-\infty}^{+\infty} f(x, y) d x d y \\
& =\int_{B} f_{Y}(y) d y
\end{aligned}
$$

## Expectation of Jointly Distributed RVs

For two discrete RVs $X$ and $Y$, we can write

$$
E[g(X, Y)]=\sum_{y} \sum_{x} g(x, y) p(x, y)
$$

For two continuous RVs $X$ and $Y$, we can write

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

## Example: Expectation of a Binomial RV

If $X$ is a binomial RV with parameters $n$ and $p$, we can write,

$$
X=\sum_{i=1}^{n} x_{i}
$$

where $X_{i}$ is a Bernoulli RV each with expectation $E(X)=p$. Hence,

$$
E(X)=\sum_{i=1}^{n} E\left(X_{i}\right)=n p
$$

## Example

Suppose that there are 25 different types of coupons and suppose that each time we select a coupon, it is equally likely to be any one of the 25 types. Compute the expected number of different types that are contained in a set of 10 coupons.

We first let $X$ be the number of different coupons in a set of 10 coupons.

$$
X=\sum_{i=1}^{25} X_{i}
$$

where

$$
X_{i}= \begin{cases}1, & \text { if at least } 1 \text { type } i \text { coupon is in the set } \\ 0, & \text { otherwise }\end{cases}
$$

## Example

We thus have,
$E\left(X_{i}\right)=P\left\{X_{i}=1\right\}$
$=P\{$ if at least 1 type $i$ coupon is in the set $\}$
$=1-P\{$ no type $i$ coupon is in the set $\}$
$=1-\left(\frac{24}{25}\right)^{10} \Rightarrow E(X)=\sum_{i=1}^{25} E\left(X_{i}\right)=25\left[1-\left(\frac{24}{25}\right)^{10}\right]$

## Independent RVs

The RVs $X$ and $Y$ are said to be independent if, for all $a, b$,

$$
P\{X \leq a, Y \leq b\}=P\{X \leq a\} P\{Y \leq b\}
$$

In terms of joint CDFs,

$$
F(a, b)=F_{X}(a) F_{Y}(b), \quad \forall a, b
$$

When $X$ and $Y$ are discrete,

$$
p(x, y)=p_{x}(x) p_{Y}(y)
$$

and when continuous

$$
f(x, y)=f_{x}(x) f_{Y}(y)
$$

## Covariance and Variance of Sums of RVs

$$
\begin{aligned}
\operatorname{Cov}(X, Y) & =E\{[X-E(X)][Y-E(Y)]\} \\
& =E[X Y-Y E(X)-X E(Y)+E(X) E(Y)] \\
& =E(X Y)-E(Y) E(X)-E(X) E(Y)+E(X) E(Y) \\
& =E(X Y)-E(X) E(Y)
\end{aligned}
$$

Note that, when $X$ and $Y$ are independent, $\operatorname{Cov}(X, Y)=0$.

## Properties of Covariance

For any RVs, $X, Y, Z$ and constant $c$, we have

- $\operatorname{Cov}(X, X)=\operatorname{Var}(X)$
- $\operatorname{Cov}(X, Y)=\operatorname{Cov}(Y, X)$
- $\operatorname{Cov}(c X, Y)=c \operatorname{Cov}(X, Y)$
- $\operatorname{Cov}(X, Y+Z)=\operatorname{Cov}(X, Y)+\operatorname{Cov}(X, Z)$

We can also show that

$$
\operatorname{Cov}\left(\sum_{i=1}^{n} X_{i}, \sum_{j=1}^{m} Y_{j}\right)=\sum_{i=1}^{n} \sum_{j=1}^{m} \operatorname{Cov}\left(X_{i}, Y_{j}\right)
$$

Variance of the Sum of RVs

$$
\begin{aligned}
\operatorname{Var}\left(\sum_{i=1}^{n} x_{i}\right) & =\operatorname{Cov}\left(\sum_{i=1}^{n} x_{i}, \sum_{j=1}^{m} x_{j}\right) \\
& =\sum_{i=1}^{n} \sum_{j=1}^{m} \operatorname{Cov}\left(X_{i}, X_{j}\right) \\
& =\sum_{i=1}^{n} \operatorname{Cov}\left(X_{i}, X_{i}\right)+\sum_{i=1}^{n} \sum_{j \neq i} \operatorname{Cov}\left(X_{i}, X_{j}\right) \\
& =\sum_{i=1}^{n} \operatorname{Var}\left(X_{i}\right)+\sum_{i=1}^{n} \sum_{j<i} \operatorname{Cov}\left(X_{i}, X_{j}\right)
\end{aligned}
$$

## Sample Mean

If $X_{1}, \ldots, X_{n}$ are IID, then, the RV $\bar{X}$ is called the sample mean and defined as

$$
\bar{X}=\frac{\sum_{i=1}^{n} X_{i}}{n}
$$

## Proposition

Suppose that $X_{1}, \ldots, X_{n}$ are IID with with expected value $\mu$ and variance $\sigma^{2}$. We can then write

- $E(\bar{X})=\mu$
- $\operatorname{Var}(\bar{X})=\sigma^{2} / n$
- $\operatorname{Cov}\left(\bar{X}, X_{i}-\bar{X}\right)=0, i=1, \ldots, n$


## Variance of a Binomial RV

We can let $X=\sum_{i=1}^{n} X_{i}$ where $X_{i}$ are independent Bernoulli RVs. Hence,

$$
\begin{aligned}
X=\sum_{i=1}^{n} X_{i} \Rightarrow \operatorname{Var}(X) & =\operatorname{Var}\left(\sum_{i=1}^{n} x_{i}\right) \\
& =\operatorname{Var}\left(\sum_{i=1}^{n}\left\{E\left(X_{i}^{2}\right)-\left[E\left(X_{i}\right)\right]^{2}\right\}\right) \\
& =\operatorname{Var}\left(\sum_{i=1}^{n}\left\{E\left(X_{i}\right)-\left[E\left(X_{i}\right)\right]^{2}\right\}\right) \\
& =\operatorname{Var} \sum_{i=1}^{n}\left(p-p^{2}\right) \\
& =n p(1-p)
\end{aligned}
$$

## Sum of 2 Independent Uniform RVs

If $X$ and $Y$ are independent uniform variables on ( 0,1 ), the probability distribution of $X+Y$,

$$
\begin{aligned}
f_{X+Y}(a)=\int_{0}^{1} f(a-y) d y \Rightarrow f_{X+Y}(a) & =\int_{0}^{a} d y=a, 0 \leq a \leq 1 \\
f_{X+Y}(a) & =\int_{a-1}^{1} d y=2-a, 1<a<2
\end{aligned}
$$

Hence,

$$
f_{X+Y}(a)=\left\{\begin{array}{cc}
a, & 0 \leq a \leq 1 \\
2-a, & 1<a<2 \\
0, & \text { otherwise }
\end{array}\right.
$$

## Sums of Independent Poisson RVs

Let $X$ and $Y$ be independent Poisson RV s with respective means $\lambda_{1}$ and $\lambda_{2}$. What is the distribution of $X+Y$ ?

The event $\{X+Y=n\}$ can be written as the intersection of the mutually exclusive events $\{X=k, Y=n-k\}$, where $0 \leq k \leq n$.

## Sums of Independent Poisson RVs

$$
\begin{aligned}
P\{X+Y=n\} & =\sum_{k=0}^{n} P\{X=k, Y=n-k\} \\
& =\sum_{k=0}^{n} P\{X=k\} P\{Y=n-k\} \\
& =\sum_{k=0}^{n} \frac{e^{-\lambda_{1}} \lambda_{1}^{k}}{k!} \frac{e^{-\lambda_{2}} \lambda_{2}^{n-k}}{(n-k)!} \\
& =e^{\lambda_{1}+\lambda_{2}} \sum_{k=0}^{n} \frac{\lambda_{1}^{k} \lambda_{2}^{n-k}}{k!(n-k)!} \\
& =\frac{e^{\lambda_{1}+\lambda_{2}}}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \lambda_{1}^{k} \lambda_{2}^{n-k} \\
& =\frac{e^{\lambda_{1}+\lambda_{2}}}{n!}\left(\lambda_{1}+\lambda_{2}\right)^{n} \Rightarrow \text { Poisson with mean } \lambda_{1}+\lambda_{2}
\end{aligned}
$$

## Moment Generating Function

The Moment Generating Function (MGF) $\phi(t)$ of RV $X$ is defined, $\forall t$, as

$$
\phi(t)=E\left(e^{t x}\right)=\sum_{x} e^{t x} p(x)
$$

when $X$ is discrete and

$$
\phi(t)=E\left(e^{t X}\right)=\int_{-\infty}^{+\infty} e^{t x} f(x) d x
$$

when $X$ is continuous.

## Moment Generating Function

All moments of $X$ can be obtained by successively differentiating the MGF.

$$
\begin{aligned}
\phi^{\prime}(t) & =\frac{d}{d t} E\left(e^{t X}\right) \\
& =E\left[\frac{d}{d t}\left(e^{t X}\right)\right] \\
& =E\left(X e^{t X}\right) \Rightarrow \phi^{\prime}(0)=E(X)
\end{aligned}
$$

## Moment Generating Function

Also note that

$$
\begin{aligned}
\phi^{\prime \prime}(t) & =\frac{d}{d t} \phi^{\prime}(t) \\
& =\frac{d}{d t} E\left(X e^{t X}\right) \\
& =E\left[\frac{d}{d t}\left(X e^{t X}\right)\right] \\
& =E\left(X^{2} e^{t X}\right) \Rightarrow \phi^{\prime \prime}(0)=E\left(X^{2}\right)
\end{aligned}
$$

## Moment Generating Function

In general, the $n$th derivative of the MGF evaluated at $t=0$ equals $E\left(X^{n}\right)$, that is,

$$
\phi^{n}(0)=E\left(X^{n}\right), \quad n \geq 1
$$

## Example

Compute the variance of a binomial RV with parameters $n, p$.

$$
\begin{aligned}
\phi(t) & =E\left(e^{t X}\right) \\
& =\sum_{k=0}^{n} e^{t k}\binom{n}{k} p^{k}(1-p)^{n-k} \\
& =\sum_{k=0}^{n}\binom{n}{k}\left(p e^{t}\right)^{k}(1-p)^{n-k} \\
& =\left(p e^{t}+1-p\right)^{n}
\end{aligned}
$$

## Example

## Since

$$
\phi(t)=\left(p e^{t}+1-p\right)^{n}
$$

the moments are

$$
\begin{aligned}
\phi^{\prime}(t) & =n\left(p e^{t}+1-p\right)^{n-1} p e^{t} \\
\phi^{\prime \prime}(t) & =n(n-1)\left(p e^{t}+1-p\right)^{n-2}\left(p e^{t}\right)^{2}+n\left(p e^{t}+1-p\right)^{n-1}
\end{aligned}
$$

and so

$$
E(X)=\phi^{\prime}(0)=n p \text { and } E\left(X^{2}\right)=\phi^{\prime \prime}(0)=n(n-1) p^{2}+n p=n p(1-p)
$$

## Example

Compute the variance of a Poisson RV with parameter $\lambda$.

$$
\begin{aligned}
\phi(t) & =E\left(e^{t X}\right) \\
& =\sum_{n=0}^{\infty} e^{t n} e^{-\lambda} \frac{\lambda^{n}}{n!} \\
& =e^{-\lambda} \sum_{n=0}^{\infty} \frac{\left(\lambda e^{t}\right)^{n}}{n!} \\
& =e^{-\lambda} \exp \left\{\lambda\left(e^{t}\right)\right\} \\
& =\exp \left\{\lambda\left(e^{t}-1\right)\right\}
\end{aligned}
$$

## Example

## Since

$$
\phi(t)=\exp \left\{\lambda\left(e^{t}-1\right)\right\}
$$

the moments are

$$
\begin{aligned}
\phi^{\prime}(t) & =\lambda e^{t} \exp \left\{\lambda\left(e^{t}-1\right)\right\} \\
\phi^{\prime \prime}(t) & =\left(\lambda e^{t}\right)^{2} \exp \left\{\left(\lambda e^{t}-1\right)\right\}+\lambda e^{t} \exp \left\{\lambda\left(e^{t}-1\right)\right\}
\end{aligned}
$$

and so

$$
E(X)=\phi^{\prime}(0)=\lambda \text { and } E\left(X^{2}\right)=\phi^{\prime \prime}(0)=\lambda^{2}+\lambda
$$

The variance is computed as

$$
\operatorname{Var}(X)=E\left(X^{2}\right)-[E(X)]^{2}=\lambda
$$

## Example

Compute the variance of an exponential RV with parameter $\lambda$.

$$
\begin{aligned}
\phi(t) & =E\left(e^{t x}\right) \\
& =\int_{0}^{\infty} e^{t x} \lambda e^{-\lambda x} d x \\
& =\lambda \int_{0}^{\infty} e^{-(\lambda-t) x} d x \\
& =\frac{\lambda}{\lambda-t} \text { for } t<\lambda
\end{aligned}
$$

## Example

We then have,

$$
\phi(t)=\frac{\lambda}{\lambda-t} \Rightarrow\left\{\begin{array}{l}
\phi^{\prime}(t)=\frac{\lambda}{(\lambda-t)^{2}} \\
\phi^{\prime \prime}(t)=\frac{2 \lambda}{(\lambda-t)^{3}}
\end{array}\right.
$$

and so

$$
E(X)=\phi^{\prime}(0)=\frac{1}{\lambda} \text { and } E\left(X^{2}\right)=\phi^{\prime \prime}(0)=\frac{2}{\lambda^{2}}
$$

The variance is computed as

$$
\operatorname{Var}(X)=E\left(X^{2}\right)-[E(X)]^{2}=\frac{1}{\lambda^{2}}
$$

## Example

Compute the variance of a normal RV with parameters $\mu, \sigma^{2}$.

$$
\begin{aligned}
\phi(t) & =E\left(e^{t z}\right) \\
& =\frac{1}{\sqrt{2 \pi}} \int_{-\infty}^{+\infty} e^{t x} e^{-x^{2} / 2} d x \\
& =\frac{1}{\sqrt{2 \pi}} \int_{-\infty}^{+\infty} e^{-\left(x^{2}-2 t x\right) / 2} d x \\
& =\frac{e^{t^{2} / 2}}{\sqrt{2 \pi}} \int_{-\infty}^{+\infty} e^{-(x-t)^{2} / 2} d x \\
& =e^{t^{2} / 2}
\end{aligned}
$$

Note that $Z$ is a standard normal RV, which means, $X=\sigma Z+\mu$ is a normal RV with parameters $\mu$ and $\sigma^{2}$.

## Example

We can then write,

$$
\begin{aligned}
& \phi(t)=E\left(e^{t X}\right)=E\left[e^{(\sigma Z+\mu)}\right]=e^{t \mu} E\left(e^{t \sigma Z}\right)=\exp \left\{\mu t+\frac{\sigma^{2} t^{2}}{2}\right\} \\
& \phi^{\prime}(t)=\left(\mu+\sigma^{2} t\right) \exp \left\{\mu t+\frac{\sigma^{2} t^{2}}{2}\right\} \\
& \phi^{\prime \prime}(t)=\left(\mu+\sigma^{2} t\right)^{2} \exp \left\{\mu t+\frac{\sigma^{2} t^{2}}{2}\right\}+\sigma^{2} \exp \left\{\mu t+\frac{\sigma^{2} t^{2}}{2}\right\}
\end{aligned}
$$

and so
$E(X)=\phi^{\prime}(0)=\mu$ and $E\left(X^{2}\right)=\phi^{\prime \prime}(0)=\mu^{2}+\sigma^{2} \Rightarrow \operatorname{Var}(X)=\sigma^{2}$

## MGF of the Sum of Independent RVs

MGF of the sum of independent $R V$ s is just the product of the individual MGFs. To show this, assume that independent RVs $X$ and $Y$ have MGFs $\phi_{X}(t)$ and $\phi_{Y}(t)$, respectively. The MGF of $X+Y$ is given as

$$
\begin{aligned}
\phi_{X+Y}(t) & =E\left[e^{t(X+Y)}\right] \\
& =E\left(e^{t X} e^{t Y}\right) \quad(\text { by independence }) \\
& =E\left(e^{t X}\right) E\left(e^{t Y}\right) \\
& =\phi_{X}(t) \phi_{Y}(t)
\end{aligned}
$$

## The Laplace Transform

For a non-negative RV $X$, it is convinient to use the Laplace transform $g(t), t \geq 0$ as

$$
g(t)=\phi(-t)=E\left(e^{-t X}\right)
$$

The advantage of using the Laplace transform is that it is always between 0 and 1 when the RV is non-negative. That is,

$$
0 \leq e^{-t X} \leq 1, X \geq 0, t \geq 0
$$

Also note that, like for MGFs, if the RVs have the same Laplace transform, they must have the same distribution (for non-negative RVs).

## The Chi-Squared Distribution

## Definition

If $Z_{1}, \ldots, Z_{n}$ are independent standard normal RVs, then, the RV $X$, defined as follows, is said to be a chi-squared RV with $n$ degrees of freedom.

$$
X=\sum_{i=1}^{n} Z_{i}^{2} \sim \mathrm{C}-\mathrm{S}(n)
$$

## The Chi-Squared Distribution

Proposition
If $X_{1}, \ldots, X_{n}$ are IID normal RV s with mean $\mu$ and variance $\sigma^{2}$, then, the sample mean $\bar{X}$ and the sample variance $S^{2}$ are independent, and $\bar{X}$ is a normal RV with mean $\mu$ and variance $\sigma^{2} / n$.
We also note that the following is a chi-squared RV with $n-1$ degrees of freedom:

$$
\frac{(n-1) S^{2}}{\sigma^{2}} \sim \operatorname{C-S}(n-1)
$$

## Limit Theorems

Proposition (Markov's Inequality)
If $X$ is a non-negative RV , then, for any $k$,

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

Proposition (Chebyshev's Inequality)
If $X$ is an RV with mean $\mu$ and variance $\sigma^{2}$, then, for any $k$,

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

## Limit Theorems

Theorem (Strong Law of Large Numbers)
Let $X_{1}, \ldots, X_{n}$ be a sequence of IID RVs, and let $E\left(X_{i}\right)=\mu$, $i=1, \ldots, n$. We can then write,

$$
\lim _{n \rightarrow \infty} \frac{\sum_{i=1}^{n} X_{i}}{n}=\mu
$$

## Limit Theorems

Theorem (Central Limit Theorem)
Let $X_{1}, \ldots, X_{n}$ be a sequence of IID RV s, and let $E\left(X_{i}\right)=\mu$ and $\operatorname{Var}\left(X_{i}\right)=\sigma^{2}, i=1, \ldots, n$. We can then write,

$$
\lim _{n \rightarrow \infty} P\left\{\frac{\sum_{i=1}^{n} X_{i}-n \mu}{\sigma \sqrt{n}} \leq k\right\}=\frac{1}{\sqrt{2 \pi}} \int_{-\infty}^{k} e^{-x^{2} / 2} d x
$$

## Limit Theorems

Normal Approximation to Binomial
If $X$ is binomial with parameters $n$ and $p$, then, we can write

$$
\lim _{n \rightarrow \infty} P\left\{\frac{X-E(X)}{\sqrt{\operatorname{Var}(X)}}=\frac{X-n p}{\sqrt{n p(1-p)}} \leq k\right\}=\frac{1}{\sqrt{2 \pi}} \int_{-\infty}^{k} e^{-x^{2} / 2} d x
$$

The normal approximation is generally quite good when $n p(1-p) \geq 10$.

## Example: Normal Approximation to Binomial

Let $X$ be the number of heads when we flip a coin for 40 times.
Find the probability that $X=20$.

$$
\begin{aligned}
P\{X=20\} & =P\{19.5<X<20.5\} \\
& =P\left\{\frac{19.5-20}{\sqrt{10}}<\frac{X-20}{\sqrt{10}}<\frac{20.5-20}{\sqrt{10}}\right\} \\
& =P\left\{-0.16<\frac{X-20}{\sqrt{10}}<0.16\right\} \\
& =\Phi(0.16)-\Phi(-0.16)=0.1272
\end{aligned}
$$

wheras the exact result is

$$
P\{X=20\}=\binom{40}{20}\left(\frac{1}{2}\right)^{20}\left(1-\frac{1}{2}\right)^{20}=0.1268
$$

## Statistical Inference and Estimation

A statistic $\hat{\theta}$ is an unbiased estimator of the population parameter $\theta$ if

$$
E(\hat{\theta})=\theta
$$

and, of all the estimators, the one with the smallest variance is called the most efficient estimator of of $\theta$.
Given independent observations $x_{1}, \ldots, x_{n}$ from a PMF or PDF, the maximum likelihood estimator (MLE) $\hat{\theta}$ maximizes the likelihood function

$$
L\left(x_{1}, \ldots, x_{n} ; \theta\right)=f\left(x_{1}, \theta\right) \ldots f\left(x_{n}, \theta\right)
$$

## Statistical Inference and Estimation

- Can you use $\bar{x}$ and $s^{2}$ to estimate $\mu$ and $\sigma^{2}$, respectively? Why?
- For example, can you use the following likelihood function to obtain the MLEfor a Poisson distribution with mean $\lambda$ ?

$$
L\left(x_{1}, \ldots, x_{n} ; \lambda\right)=\prod_{i=1}^{n} p\left(x_{i} \mid \lambda\right)=\frac{e^{-n \lambda} \lambda\left(\sum_{i=1}^{n} x_{i}\right)}{\prod_{i=1}^{n}\left(x_{i}\right)!}
$$

The log-likelihood is then

$$
\begin{aligned}
\ln \left[L\left(x_{1}, \ldots, x_{n} ; \lambda\right)\right] & =-n \lambda+\sum_{i=1}^{n} \ln \lambda-\ln \left(\prod_{i=1}^{n}\left(x_{i}\right)!\right) \\
\frac{\partial \ln \left[L\left(x_{1}, \ldots, x_{n} ; \lambda\right)\right]}{\partial \lambda} & =-n+\sum_{i=1}^{n} \frac{x_{i}}{\lambda} \quad(\text { continue } \ldots)
\end{aligned}
$$

## Stochastic Processes

- A stochastic process $\{X(t), t \in T\}$ is a collection of RV .
- The index $t$ is often referred as time.
- We refer to $X(t)$ as the state of the process at time $t$.
- The set $T$ is called as the index set.
- When $T$ is countable, we refer to the stochastic process as the discrete-time process.
- When $T$ is uncountable, we refer to the stochastic process as the continuous-time process.

The End
Questions?

