Now for some continuous RV examples.

**Uniform RV**

\( Y \sim \text{Unif}(a,b) \) - \( Y \) is uniformly distributed over the interval \((a,b)\)

\( \Rightarrow \)

\( f_Y(y) = \begin{cases} \frac{1}{b-a} & \text{for } a < y < b \\ 0 & \text{o.w.} \end{cases} \)

PDF of \( Y \): \[
\frac{1}{\text{length of interval}}
\]

**Exponential RV**

\( Y \sim \text{exp}(\lambda) \) - exponential with rate \( \lambda \)

PDF: \[
f_Y(y) = \lambda e^{-\lambda y}, \quad y \geq 0
\]

\( \Rightarrow \)

Continuous version of geometric distribution

Time duration until an event

**Gamma RV**

Cont. version of Neg. Binomial: Time until \( r \)th event
- Gamma ($\alpha=1$, $\beta$) = exponential ($\beta$)
- Sum of $r$ exponential ($\beta$) RVs is gamma($r$, $\beta$)

Normal RV

$X \sim N(\mu, \sigma^2)$

\[
\text{PDF: } f_X(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{for } x \in \mathbb{R}
\]

[Sneak Peak]: Bernoulli Process

- Sequence of indep. and identically distributed Bernoulli trials (RVs)
  
  $X_1, X_2, X_3, X_4, X_5, \ldots$ where $X_i \sim \text{Bernoulli}(p)$

  e.g. 0 1 1 0 1 0 ...
  realization

- Discrete-time stochastic process that only takes 2 values: 0 or 1

  repeated coin flips via a possibly unfair coin (consistent unfairness) → see print out for more details
The two possible values of each $X_i$ are often called "success" and "failure". Thus, when expressed as a number 0 or 1, the outcome may be called the number of successes on the $i$th "trial".

Two other common interpretations of the values are true or false and yes or no. Under any interpretation of the two values, the individual variables $X_i$ may be called Bernoulli trials with parameter $p$.

In many applications time passes between trials, as the index $i$ increases. In effect, the trials $X_1, X_2, \ldots, X_i, \ldots$ happen at "points in time" 1, 2, \ldots, $i$, \ldots That passage of time and the associated notions of "past" and "future" are not necessary, however. Most generally, any $X_i$ and $X_j$ in the process are simply two from a set of random variables indexed by $\{1, 2, \ldots, n\}$ or by $\{1, 2, 3, \ldots\}$, the finite and infinite cases.

Several random variables and probability distributions beside the Bernoullis may be derived from the Bernoulli process:

- The number of successes in the first $n$ trials, which has a binomial distribution $B(n, p)$
- The number of trials needed to get $r$ successes, which has a negative binomial distribution $NB(r, p)$
- The number of trials needed to get one success, which has a geometric distribution $NB(1, p)$, a special case of the negative binomial distribution

The negative binomial variables may be interpreted as random waiting times.
Expectation & Variance

**def**: The expected value of RV $X$ is

$$E[X] = \begin{cases} \sum \limits_{k} k \cdot P_X(k) = \sum \limits_{k} k \cdot P(X=k), & \text{if } X \text{ discrete} \\ \int_{-\infty}^{\infty} x \cdot f_X(x) dx, & \text{if } X \text{ continuous} \end{cases}$$

**def**: The variance of RV $X$ (which has mean $\mu = E[X]$) is

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

**def**: The standard deviation of $X$ is $\sigma = \sqrt{\text{Var}(X)}$.

**Properties**

- $E[aX+b] = a \cdot E[X] + b$ (linearity)
- $\text{Var}(aX+b) = a^2 \cdot \text{Var}(X)$

**Joint Distributions**

Now consider 2 RVs $X \in Y$ (generalize to an arbitrary # of RVs).

The joint distribution function (CDF) of $X \in Y$:

- **Discrete**: $F_{X,Y}(x,y) = P(X=x, Y=y) = \sum_{j \leq x, k \leq y} P(X=j, Y=k)$
- **Continuous**: $F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v) du dv$
Independent RVs

\( X \perp Y \) are independent if

\[
P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)
\]

\[F_{X,Y}(x,y) = F_X(x)F_Y(y)\]

Similarly with PDFs:

\[
P(X=x,Y=y) = P(X=x)P(Y=y)
\]

\[f_{X,Y}(x,y) = f_X(x)f_Y(y)\]

(continuous case)

\[
P(X=x,Y=y) = P(X=x)P(Y=y)
\]

\[f_{X,Y}(x,y) = f_X(x)f_Y(y)\]

(discrete case)

If \( X \perp Y \) are independent, then

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

i.e. \( E[XY] = E[X]E[Y] \)

* * *

If want more review notes, I'll post Exam 2 review from Prob. Course

Moment Generating Functions (MGF)

\[\phi(t) = E[e^{tX}] = \begin{cases}
\sum_x e^{tx}p_X(x), & \text{if } X \text{ discrete} \\
\int_{-\infty}^{\infty} e^{tx}f_X(x)dx, & \text{if } X \text{ continuous}
\end{cases}\]

\[\phi(0) = 1\]

* Called MGF b/c all moments of \( X \) can be obtained by successively differentiating \( \phi(X) \).

\[\phi'(t) = \frac{d}{dt} E[e^{tX}] = E\left[\frac{d}{dt} e^{tX}\right] = E[Xe^{tX}]\]

\[\Rightarrow \phi'(0) = E[X] .\]
Also, $\phi''(0) = E[X^2]$. In general,

$$\phi^n(0) = E[X^n], \quad n \geq 1$$

$n^{th}$ derivative w.r.t. $t$ evaluated at $t=0$

$n^{th}$ moment of $X$

(Uniqueness)

Key Property: If 2 random variables have the same MGF, then they have the same distribution (same PDF, CDF).

Independence & MGFs: Let $X = X_1 + \cdots + X_n$ where $X_i$'s are indep. RVs. Then

$$\phi_X(t) = \phi_{X_1 + \cdots + X_n}(t) = \phi_{X_1}(t) \cdot \cdots \cdot \phi_{X_n}(t)$$

Limit Theorems

Start with 2 inequalities that are useful for deriving bounds on probabilities when only the mean (or both mean & variance) of the prob. dist'n is known.

Markov Inequality

If $X$ is a RV s.t. $P(X \geq 0) = 1$, then

$$P(X \geq t) \leq \frac{E[X]}{t} \quad \text{for } t > 0$$

\text{nonneg. RV}
Chebyshev's Inequality

If \( X \) is a RV with finite variance \( \sigma^2 \), then

\[
P\left( |X - E[X]| \geq t \right) \leq \frac{\sigma^2}{t^2} \quad \text{for } t > 0
\]

Special Cases: Let \( t = k\sigma \), \( \frac{1}{4} E[X] = \mu \)

\[
P\left( |X - \mu| \geq k\sigma \right) \leq \frac{1}{k^2} \quad \text{OR} \quad P\left( |X - \mu| < k\sigma \right) \geq 1 - \frac{1}{k^2}
\]

Law of Large Numbers (LLN) - Strong Law

Let \( X_1, X_2, \ldots \) be a sequence of independent RVs with the same distribution (i.i.d RVs), and let \( \mu = E[X_i] \).

Then, with probability 1,

\[
\lim_{n \to \infty} \frac{X_1 + \cdots + X_n}{n} = \mu
\]

(simply,
\[
\bar{X}_n \xrightarrow{a.s.} \mu \text{ as } n \to \infty
\]

where,
\[
\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i
\]

\[
\left( i.e., \ P\left( \lim_{n \to \infty} \bar{X}_n = \mu \right) = 1 \right)
\]

[Weak LLN is similar, except \( \bar{X}_n \xrightarrow{P} \mu \)]

\[
\text{Sample mean converges to theoretical mean } \mu
\]

[i.e., convergence in Probability]

\[
\lim_{n \to \infty} P\left( |\bar{X}_n - \mu| \leq \varepsilon \right) = 1
\]
Central Limit Theorem (holds for any distribution!)

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. RVs with finite mean $\mu$ and variance $\sigma^2$. Then

$$
\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \xrightarrow{D} Z \sim \mathcal{N}(0,1)
$$

"converges in distribution"  \hspace{1cm} \text{standard normal dist'n}

i.e. $P\left(\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \leq a\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} \, dx = P(Z \leq a)$

\[
\begin{align*}
\text{Recall: } & \quad \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i, & E[\bar{X}_n] &= \mu, & \text{Var}(\bar{X}_n) &= \frac{\sigma^2}{n} \\
& \text{Sample mean} & & & & \\
\end{align*}
\]

* Main Idea of CLT *

Sample mean $\bar{X}_n$ is approximately normally distributed of a sufficiently large # of iid RVs with mean $\mu$ & var $\sigma^2/n$ regardless of underlying distribution.

i.e. $P\left(\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \leq a\right) \approx P(Z \leq a)$
R practice w/ CLT & other applications from today's lecture!

2 R scripts on website

LLN

CLT

(work through these 2nd half of class)
Convergence Concepts $\frac{1}{2}$ LLN

(Also see Paul’s handout)

def: A sequence of RVs $X_n$ converges in distribution to RV $X$ if

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

for all $x$ where $F_X(x)$ is continuous. This is pointwise convergence of CDFs.

def: A seq. of RVs $X_n$ converges in probability to RV $X$ if $\forall \varepsilon > 0$,

$$\lim_{n \to \infty} P(\left|X_n - X\right| \geq \varepsilon) = 0$$

$$\iff \lim_{n \to \infty} P(\left|X_n - X\right| \leq \varepsilon) = 1$$

def: A seq. of RVs $X_n$ converges almost surely to RV $X$ if $\forall \varepsilon > 0$,

$$P\left(\lim_{n \to \infty} \left|X_n - X\right| \leq \varepsilon\right) = 1$$

$$\left(P\left(\forall \omega \in \Omega \text{ s.t. } \lim_{n \to \infty} \left|X_n(\omega) - X(\omega)\right| \leq \varepsilon\right) = 1\right)$$

Measure: “almost everywhere” means a stunt holds true for all but a set of
Thinking of RVs as functions on our sample space $S$, this is just pointwise convergence of the RVs except perhaps on some set of measure 0.

**Thm:** If $X_n$ converges a.s. to $X$, then it also converges in probability to $X$.

If $X_n$ converges in prob. to $X$, then it also converges in distribution to $X$.

\[ X_n \xrightarrow{\text{a.s.}} X \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{D} X \]

**Weak Law of Large Numbers (WLLN)**

Let $X_i$ be i.i.d. RVs with mean $\mu$. Then $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ converges in probability to $\mu$, i.e.,

\[ \bar{X}_n \xrightarrow{P} \mu. \]

\[ (\forall \varepsilon > 0 \text{ near zero } \lim_{n \to \infty} P(|\bar{X}_n - \mu| \leq \varepsilon) = 1) \]

**Strong Law of Large Numbers (SLLN)**

(Same setup as WLLN) Then

\[ \bar{X}_n \text{ converges almost surely to } \mu, \text{ i.e. } \bar{X}_n \xrightarrow{\text{a.s.}} \mu. \]

\[ (\forall \varepsilon > 0, \ P(\lim_{n \to \infty} |\bar{X}_n - \mu| \leq \varepsilon) = 1) \]