Now for some continuous RV examples.

Uniform RV

$Y \sim Unif(a,b)$ - Y is uniformly distributed over the interval (a,b)
("All outcomes are equally likely")

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

$\frac{1}{\text{length of interval}}$

Exponential RV

$Y \sim exp(\lambda)$ - exponential with rate

PDF: $\boxed{f_Y(y) = \lambda e^{-\lambda y}, \ y \geq 0}$ (& 0 otherwise)

... Continuous version of geometric distribution

Time duration until an event

Gamma RV

Cont. version of Neg. Binomial : Time until r^{th} event

- $\text{Gamma}(\overset{\text{shape}}{\alpha = 1}, \overset{\text{scale}}{\beta}) = \text{exponential}(\beta)$

- Sum of r exponential (β) RVs is gamma(r, β)

Normal RV

$$X \sim N(\underset{\text{mean}}{\mu}, \underset{\text{var}}{\sigma^2})$$

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

$x \in (-\infty, \infty)$

[Sneak Peak on Stochastic Processes] : Bernoulli Process

- Sequence of indep. & 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. realization: 0 1 1 0 1 ...

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

... repeated coin flips w/ 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 ith "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 (mean) value of RV X is

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

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

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

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

Properties

- $E[aX+b] = a\,E[X] + b$ (linearity)
- $Var(aX+b) = a^2\,Var(X)$

Joint Distributions

Now consider 2 RVs X & Y (can generalize to an arbitrary # of RVs).

The joint distribution function (CDF) of X & Y:

$$F_{X,Y}(x,y) = P(X \le x, Y \le y) = \begin{cases} \sum_{j \le x} \sum_{k \le y} P(X=j, Y=k) & \text{discrete} \\ \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)\,du\,dv & \text{continuous} \end{cases}$$

for $x, y \in \mathbb{R}$

Independent RVs

X & Y are independent if $P(X \leq x, Y \leq y) = P(X \leq x) P(Y \leq y)$

joint CDF is product of marginal CDFs $\boxed{F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)}$

Similarly with PDFs:

$$\boxed{P(X = x, Y = y) = P(X = x) P(Y = y)} \quad \text{discrete case}$$

$$\boxed{f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)} \quad \text{continuous case}$$

If X & Y are independent, then

$$\boxed{E[g(X) h(Y)] = E[g(X)] E[h(Y)]}$$

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)

def: The moment generating function of RV X is

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

for all $t \in \mathbb{R}$ for which the expected value exists.

Note: $\phi(0) = 1$

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

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

$\Rightarrow \phi'(0) = E[X]$.

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

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

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

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

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   nonneg. RV

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

Chebyshev's Inequality

If X is a RV with finite variance σ^2, then

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

Special Cases: Let $t = k\sigma$ & $E[X] = \mu$

$$\boxed{P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}} \quad \text{OR} \quad \boxed{P(|X-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}}$$

stopped

Law of Large Numbers (LLN) - Strong Law

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

Then, with probability 1,

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mu \qquad \left(\text{Simply, } \bar{X}_n \xrightarrow{a.s.} \mu \text{ as } n \to \infty \right)$$

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

where $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

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

convergence in Probability

i.e. $\lim_{n\to\infty} P(|\bar{X}_n - \mu| \leq \varepsilon) =$

Sample mean converges to theoretical mean μ

Central Limit Theorem (CLT) (holds for any distribution!)

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

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

"converges in distribution" — standard Normal dist'n

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

Recall: $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ (sample mean), $E[\bar{X}_n] = \mu$, $Var(\bar{X}_n) = \frac{\sigma^2}{n}$

* Main Idea of CLT *

Sample mean of a sufficiently large # of iid RVs is approximately normally distributed with mean μ & var σ^2/n Regardless of underlying distribution.

i.e. $$P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le a\right) \approx P(Z \le a)$$

R practice w/ CLT today ~~on Thurs~~ & other applications from today's lecture!

2 R scripts on website

LLN

CLT

(Work through these 2nd half of class

Convergence Concepts & LLN

(Also see Paul's handout)

<u>def</u>: A sequence of i.i.d. RVs X_n <u>converges in distribution</u> to RV X if $\left(X_n \xrightarrow{D} X \right)$

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

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

<u>def</u>: A seq. of i.i.d. RVs X_n <u>converges in probability</u> to RV X if $\forall\, \varepsilon > 0$, $\left(X_n \xrightarrow{P} X \right)$

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

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

<u>def</u>: A seq. of i.i.d. RVs X_n <u>converges almost surely</u> to RV X if $\forall\, \varepsilon > 0$, $\left(X_n \xrightarrow{a.s.} X \right)$

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

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

Measure Theory: "almost everywhere" means a stmt 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.

$$\boxed{X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{P} X \implies X_n \xrightarrow{D} X}$$

Weak Law of Large Numbers (WLLN)

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

$$\left(\forall \varepsilon > 0 \text{ near zero } \lim_{n\to\infty} P\left(|\bar{X}_n - \mu| \le \varepsilon\right) = 1 \right)$$

Strong Law of Large Numbers (SLLN)

(Same setup as WLLN) Then

$\bar{X}_n$ converges almost surely to μ, i.e. $\boxed{\bar{X}_n \xrightarrow{a.s.} \mu}$.

as $n\to\infty$

$$\left(\forall \varepsilon > 0, \ P\left(\lim_{n\to\infty} |\bar{X}_n - \mu| \le \varepsilon\right) = 1 \right)$$