6.041 Fall08 Quiz II Review

1 Probability Density Functions (PDF)

For a continuous RV X with PDF $f_X(x)$ (≥ 0),

$$\begin{aligned} \mathbf{P}(a \leq X \leq b) &= \int_a^b f_X(x)dx \\ \mathbf{P}(x \leq X \leq x + \delta) &\approx f_X(x).\delta \\ \mathbf{P}(X \in A) &= \int_A f_X(x)dx \end{aligned}$$

Remarks:

- if X is continuous, $\mathbf{P}(X = x) = 0 \ \ \forall x$!!
- $f_X(x)$ may take values larger than 1.

Normalization property:

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

2 Mean and variance of a continuous RV

$$
\begin{aligned}
\mathbf{E}[X] &= \int_{-\infty}^{\infty} x f_X(x)dx \\
\mathbf{E}[g(X)] &= \int_{-\infty}^{\infty} g(x) f_X(x)dx \\
\mathrm{Var}(X) &= \int_{-\infty}^{\infty} (x - \mathbf{E}[X])^2 f_X(x)dx \\
&= \mathbf{E}[X^2] - (\mathbf{E}[X])^2 \ (\geq 0) \\
\mathbf{E}[aX + b] &= a\mathbf{E}[X] + b \\
\mathrm{Var}(aX + b) &= a^2 \mathrm{Var}(X)
\end{aligned}
$$

3 Cumulative Distribution Functions

Definition:

$$F_X(x) = \mathbf{P}(X \le x)$$

monotonically increasing from 0 (at $-\infty$) to 1 (at $+\infty$).

- Continuous RV:

$$F_X(x) = \mathbf{P}(X \le x) = \int_{-\infty}^{x} f_X(t)dt \quad \text{(continuous)}$$

$$f_X(x) = \frac{dF_X}{dx}(x)$$

- Discrete RV:

$$F_X(x) = \mathbf{P}(X \le x) = \sum_{k \le x} p_X(k) \quad \text{(piecewise constant)}$$

$$p_X(k) = F_X(k) - F_X(k-1)$$

4 Normal/Gaussian Random Variables

Standard Normal RV: $N(0,1)$:

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \\ \mathbf{E}[X] &= 0, \quad \mathrm{Var}(X) = 1 \end{aligned}$$

General normal RV: $N(\mu, \sigma^2)$:

$$\begin{aligned} f_X(x) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \\ \mathbf{E}[X] &= \mu, \quad \mathrm{Var}(X) = \sigma^2 \end{aligned}$$

- if $Y = aX + b$, then $Y \sim N(a\mu + b,\ a^2\sigma^2)$.
- CDF for standard normal $\phi(.)$ can be read in a table.
- To evaluate CDF of a general standard normal, express it as a function of a standard normal:

$$X \sim N(\mu, \sigma^2) \Leftrightarrow \frac{X - \mu}{\sigma} \sim N(0, 1)$$

$$\mathbf{P}(X \leq x) = P\Big(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\Big) = \phi\Big(\frac{x - \mu}{\sigma}\Big)$$

where $\phi(.)$ denotes the CDF of a standard normal.

5 Joint PDF

Joint PDF of two continuous RV X and Y: $f_{X,Y}(x,y)$.

$$\mathbf{P}(x \leq X \leq x+\delta,\ y \leq Y \leq y+\delta) \approx f_{X,Y}(x,y).\delta^2$$

$$\mathbf{P}(A) = \int\int_A f_{X,Y}(x,y)dxdy$$

$$\mathbf{E}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dxdy$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

By definition,

$$X,\ Y \text{ independent} \Leftrightarrow f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

6 Conditioning on an event

X a continuous RV, A a subset of the real line

$$
\begin{aligned}
f_{X|A}(x) &= \begin{cases} \frac{f_X(x)}{\mathbf{P}(X \in A)} & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases} \\
\mathbf{P}(X \in B | X \in A) &= \int_B f_{X|A}(x) dx \\
\mathbf{E}[X|A] &= \int_{-\infty}^{\infty} x f_{X|A}(x) dx \\
\mathbf{E}[g(X)|A] &= \int_{-\infty}^{\infty} g(x) f_{X|A}(x) dx
\end{aligned}
$$

If $A_1, \ldots, A_n$ are disjoint events that form a partition of the sample space,

$$f_X(x) = \sum_{i=1}^{n} \mathbf{P}(A_i) f_{X|A_i}(x) \text{ (}\approx\text{ total probability theorem)}$$

$$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{P}(A_i) \mathbf{E}[X|A_i] \text{ (total expectation theorem)}$$

$$\mathbf{E}[g(X)] = \sum_{i=1}^{n} \mathbf{P}(A_i) \mathbf{E}[g(X)|A_i]$$

7 Conditioning on a RV

X, Y continuous RV, A an event.

$$\mathbf{P}(x \leq X \leq x + \delta | Y \approx y) \approx f_{X|Y}(x|y).\delta$$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

$$f_X(x) = \int_{-\infty}^{\infty} f_Y(y) f_{X|Y}(x|y) dy$$

$$\mathbf{P}(A) = \int_{-\infty}^{\infty} \mathbf{P}(A|X = x) f_X(x) dx$$

$$
\begin{aligned}
\mathbf{E}[Y|X=x] &= \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy \\
\mathbf{E}[g(Y)|X=x] &= \int_{-\infty}^{\infty} g(y) f_{Y|X}(y|x)dy \\
\mathbf{E}[g(X,Y)|X=x] &= \int_{-\infty}^{\infty} g(x,y) f_{Y|X}(y|x)dy \\
\mathbf{E}[Y] &= \int_{-\infty}^{\infty} \mathbf{E}[Y|X=x] f_X(x)dx \\
\mathbf{E}[g(Y)] &= \int_{-\infty}^{\infty} \mathbf{E}[g(Y)|X=x] f_X(x)dx \\
\mathbf{E}[g(X,Y)] &= \int_{-\infty}^{\infty} \mathbf{E}[g(X,Y)|X=x] f_X(x)dx
\end{aligned}
$$

8 Continuous Bayes' Rule

X, Y continuous RV, N discrete RV, A an event.

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)} = \frac{f_{Y|X}(y|x) f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|t) f_X(t) dt}$$

$$\mathbf{P}(A|Y=y) = \frac{\mathbf{P}(A) f_{Y|A}(y)}{f_Y(y)} = \frac{\mathbf{P}(A) f_{Y|A}(y)}{f_{Y|A}(y)\mathbf{P}(A) + f_{Y|A^c}(y)\mathbf{P}(A^c)}$$

$$\mathbf{P}(N=n|Y=y) = \frac{p_N(n) f_{Y|N}(y|n)}{f_Y(y)} = \frac{p_N(n) f_{Y|N}(y|n)}{\sum_i p_N(i) f_{Y|N}(y|i)}$$

9 Independence of continuous RV

$$
\begin{aligned}
X,\ Y \text{ independent} \quad &\Leftrightarrow \quad f_{X|Y}(x|y) = f_X(x) \\
&\Rightarrow \quad g(X),\ h(Y) \text{ independent} \\
&\Rightarrow \quad \mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y] \\
&\Rightarrow \quad \mathbf{E}[g(X)h(Y)] = \mathbf{E}[g(X)]\mathbf{E}[h(Y)] \\
&\Rightarrow \quad \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)
\end{aligned}
$$

10 Derived distributions

Def: PDF of a *function* of a RV X with known PDF: $Y = g(X)$.
Method:

- Get the CDF:

$$F_Y(y) = \mathbf{P}(Y \leq y) = \mathbf{P}(g(X) \leq y) = \int_{x|g(x) \leq y} f_X(x) dx$$

- Differentiate: $f_Y(y) = \frac{dF_Y}{dy}(y)$

11 Convolution

$W = X + Y$, with X, Y independent.

- Discrete case:

$$p_W(w) = \sum_x p_X(x) p_Y(w - x)$$

- Continuous case:

$$f_W(w) = \int_{-\infty}^{\infty} f_X(x) f_Y(w - x)\, dx$$

Mechanics:

- put the PMFs (or PDFs) on top of each other
- flip the PMF (or PDF) of Y
- shift the flipped PMF (or PDF) of Y by w
- cross-multiply and add (or evaluate the integral)

In particular, if X, Y are independent and normal, then

- $W = X + Y$ is normal
- $f_{X|W}(x|w)$ is a normal PDF for any given w.

12 Law of iterated expectations

$\mathbf{E}[X|Y]$ is a random variable that is a function of Y
(the expectation is taken with respect to X).
To compute $\mathbf{E}[X|Y]$, first express $\mathbf{E}[X|Y=y]$ as a function of y.

Law of iterated expectations:

$$\mathbf{E}[X] = \mathbf{E}[\mathbf{E}[X|Y]]$$

(equality between two real numbers)

13 Law of conditional variances

$\text{Var}(X|Y)$ is a random variable that is a function of Y (the variance is taken with respect to X).
To compute $\text{Var}(X|Y)$, first express

$$\text{Var}(X|Y=y) = \mathbf{E}[(X - \mathbf{E}[X|Y=y])^2|Y=y]$$

as a function of y.

Law of conditional variances:

$$\text{Var}(X) = \mathbf{E}[\text{Var}(X|Y)] + \text{Var}(\mathbf{E}[X|Y])$$

(equality between two real numbers)

14 Sum of a random number of iid RVs

N discrete RV, X_i i.i.d and independent of N.
$Y = X_1 + \ldots + X_N$. Then:

$$\begin{aligned} \mathbf{E}[Y] &= \mathbf{E}[X]\mathbf{E}[N] \\ \mathrm{Var}(Y) &= \mathbf{E}[N]\mathrm{Var}(X) + (\mathbf{E}[X])^2\mathrm{Var}(N) \end{aligned}$$

15 Covariance and Correlation

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbf{E}[(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])] \\ &= \mathbf{E}[XY]-\mathbf{E}[X]\mathbf{E}[Y]\end{aligned}$$

If the covariance is positive, the random variables vary together; if it is negative, they vary inversely.

Correlation: (has no dimension)

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} \quad \in [-1,1]$$

By definition, X, Y are uncorrelated if and only if $\mathrm{Cov}(X,Y) = 0$.
Remark: X, Y independent $\Rightarrow$ $\mathrm{Cov}(X,Y) = 0$
(the converse is not true)

16 Uniform continuous RV over $[a, b]$

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{x-a}{b-a} & \text{if } a \leq x \leq b \\ 1 & \text{otherwise } (x > b) \end{cases}$$

$$\mathbf{E}[X] = \frac{a+b}{2} \qquad \text{Var}(X) = \frac{(b-a)^2}{12}$$

17 Exponential RV with parameter λ

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{E}[X] = \frac{1}{\lambda} \quad \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

18 Normal RV with parameters (μ, σ^2)

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

$$\mathbf{E}[X] = \mu$$

$$\mathrm{Var}(X) = \sigma^2$$

Limit Theorems (1)

- **Markov Inequality** If X is a **nonnegative** random variable with mean μ, then
$\mathbf{P}(X \geq a) \leq \frac{\mu}{a}$.

- **Chebyshev Inequality** If X is a random variable with mean μ and variance σ^2, then
$\mathbf{P}(\mid X - \mu \mid \geq c) \leq \frac{\sigma^2}{c^2}$, for positive c.

- **Convergence in Probability** Let Y_1, Y_2, ... be a sequence of random variables (not necessarily independent), and let a be a real number. We say that the sequence Y_n converges to a in probability, if for every $\epsilon > 0$, we have $\mathbf{P}(|Y_n - a| \geq \epsilon) \to 0$. Intuitively, as n grows large, the random variable Y_n takes on the value a with probability 1.

Limit Theorems (2)

- Central Limit Theorem: Let X_1, X_2, ... be a sequence of i.i.d. random variables with mean μ and variance σ^2. We define the normalized sum $Z_n = \frac{X_1+X_2+\ldots+X_n-n\mu}{\sigma\sqrt{n}}$
The CDF of Z_n converges to the standard normal CDF pointwise, i.e. $\mathbf{P}(Z_n \leq z) \to \Phi(z)$ for every z.

- Normal approximation based on CLT: If you have the sum of n i.i.d. random variables, to compute its CDF (and hence some probability), first subtract the mean and divide by the standard deviation and treat what you get as a standard normal random variable. This approximation gets better with larger n.

- If each X_i is Bernoulli, then approximate $\mathbf{P}(Z_n > a)$ as $\mathbf{P}(Z_n \geq a + 0.5)$.

Bernoulli Process

Bernoulli process is a sequence $X_1, X_2 \ldots$ of **independent** Bernoulli random variables with

$$\begin{aligned} \mathbf{P}(X_i = 1) &= p \\ \mathbf{P}(X_i = 0) &= 1 - p \end{aligned}$$

Memoryless property

For any given time n, the sequence $X_{n+1}, X_{n+2} \ldots$ is also a Bernoulli process, and is **independent** from $X_1, X_2 \ldots X_n$.

Fresh-Start

Every arrival restarts the process.

Important RV associated with Bernoulli Processes

- **First arrival** : The time to first arrival (T) is a **geometric** RV

$$p_T(t) = (1-p)^{t-1}p, t = 1, 2\dots$$

- **Number of arrivals**: The number of arrivals (K) in n trials is a **binomial** RV

$$p_K(k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1 \dots n$$

Note: (n-fixed,k-random)

- K^{th} **arrival**: The time to the K^{th} arrival Y_K is a **Pascal** RV

$$p_{Y_K}(t) = \binom{t-1}{k-1} p^k (1-p)^{(t-k)}$$

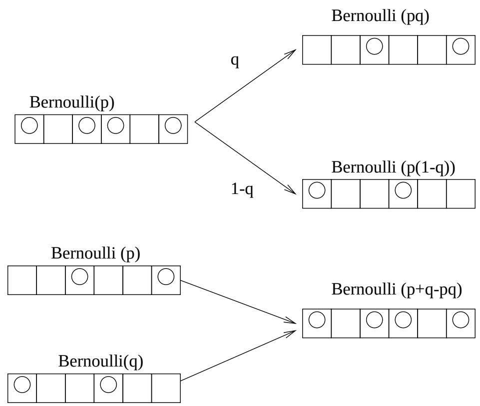
Note: (k-fixed,t-random)

Alternate description of the Bernoulli Process

- Start with a sequence of **independent geometric** RVs $T_1, T_2 \ldots$, with common parameter p.
- Record success(arrival) at times, $T_1, T_1 + T_2, T_1 + T_2 + T_3 \ldots$.
- K^{th} arrival time Y_k is the sum of the first k inter-arrival times

$$
\begin{aligned}
Y_k &= T_1 + T_2 \ldots T_k \\
\mathbf{E}[Y_k] &= \mathbf{E}[T_1 + T_2 \ldots T_k] = \frac{k}{p} \\
\mathrm{var}(Y_k) &= \mathrm{var}(T_1 + T_2 \ldots T_k) = \frac{k(1-p)}{p^2}
\end{aligned}
$$

Splitting and Merging of Bernoulli Processes



- If arrivals from a Bernoulli process are split into two processes with probability q and (1-q), each process is a Bernoulli process with parameters pq and p(1-q) respectively.
- Conversely, if we merge two indpendent Bernoulli processes with parameters p and q, we get a Bernoulli process with parameter (p+q-pq).

Summary

	Bernoulli
Inter-arrival time	Geometric
Number of arrivals	Binomial
K^{th} arrival	Pascal

Stuck on a problem?

1. Draw the graphs / trees / slots / etc.
   (a) Shade the area (if applicable) and test it with some values to make sure that area really is the event you're looking for. (Maybe you really wanted that top (or bottom, or other) area instead?)
2. Try some simple examples ($n = 1$, or 2, or 3, ...)
3. Go over the list of topics, and see if it applies to the problem.
   (a) Can this joint distribution be simplified with independence?
   (b) Would conditioning / Bayes help?
   (c) Can we use Iterated Expectations? ...

Check your answers

1. Variances should be non negative! (standard deviations should be real!)
2. Does your distribution sum or integrate to 1?
3. Check your units!
4. As certain quantities get big, does your answer do what you think it should?

Integration

- Make sure you really even need to use integration!
  - Can you use independence, or is there another simpler expectation computation?
- Try drawing the graph, then shade the area/volume you're trying to integrate over.
- Check your limits!
- Think about integration methods:
  - Substitution: make sure you change your limits to the new variable when you substitute!
  - By Parts

Dealing with mins and maxes

$$\mathbf{P}(\min\{A, B\} < k) = 1 - \mathbf{P}(\min\{A, B\} > k)$$

$$\mathbf{P}(\min\{A, B\} > k) = \mathbf{P}(A > k \cap B > k)$$

$$\mathbf{P}(\max\{A, B\} > k) = 1 - \mathbf{P}(\max\{A, B\} < k)$$

$$\mathbf{P}(\max\{A, B\} < k) = \mathbf{P}(A < k \cap B < k)$$

Usually, you can use independence to break up the intersection probability into a product.