6.041/6.431 Quiz II Review (Spring 2011)
Probability Density Functions (PDF)

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

$$P(a \leq X \leq b) = \int_a^b f_X(x) \, dx$$

$$P(x \leq X \leq x + \delta) \approx f_X(x) \cdot \delta$$

$$P(X \in A) = \int_A f_X(x) \, dx$$

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

Normalization property:

$$\int_{-\infty}^{\infty} f_X(x) \, dx = 1$$
Mean and variance of a continuous RV

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

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

\[ \text{Var}(X) = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x) \, dx \]

\[ = E[X^2] - (E[X])^2 \quad (\geq 0) \]

\[ E[aX + b] = aE[X] + b \]

\[ \text{Var}(aX + b) = a^2 \text{Var}(X) \]
Cumulative Distribution Functions

Definition:

\[ F_X(x) = P(X \leq x) \]

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

- Continuous RV:

\[ F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt \quad \text{(continuous)} \]

\[ f_X(x) = \frac{dF_X}{dx}(x) \]

- Discrete RV:

\[ F_X(x) = P(X \leq x) = \sum_{k\leq x} p_X(k) \quad \text{(piecewise constant-steps to 1)} \]

\[ p_X(k) = F_X(k) - F_X(k - 1) \quad \text{(height of step at k)} \]
Normal/Gaussian Random Variables

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

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

$$E[X] = 0, \quad \text{Var}(X) = 1$$

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

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

$$E[X] = \mu, \quad \text{Var}(X) = \sigma^2$$
• 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:

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

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

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

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

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

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

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

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

By definition,

$X, \ Y$ independent $\Leftrightarrow f_{X,Y}(x, y) = f_X(x)f_Y(y)$
Conditioning on an event

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

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{P(X \in A)} & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

$$P(X \in B|X \in A) = \int_B f_{X|A}(x)dx$$

$$E[X|A] = \int_{-\infty}^{\infty} x f_{X|A}(x)dx$$

$$E[g(X)|A] = \int_{-\infty}^{\infty} g(x) f_{X|A}(x)dx$$
If $A_1, \ldots, A_n$ are disjoint events that form a partition of the sample space,

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

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

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

$X, Y$ continuous RV, $A$ an event.

\[ P(x \leq X \leq x + \delta | Y \approx y) \approx f_{X|Y}(x|y) \cdot \delta \]

\[ f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \quad \text{(defined where denominator is positive)} \]

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

\[ P(A) = \int_{-\infty}^{\infty} P(A|X = x) f_X(x) dx \quad (\approx \text{total probability theorem}) \]
\[ E[Y | X = x] = \int_{-\infty}^{\infty} y f_{Y|x}(y|x) dy \]

\[ E[g(Y) | X = x] = \int_{-\infty}^{\infty} g(y) f_{Y|x}(y|x) dy \]

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

Versions of Total Expectation Theorem:

\[ E[Y] = \int_{-\infty}^{\infty} E[Y | X = x] f_X(x) dx \]

\[ E[g(Y)] = \int_{-\infty}^{\infty} E[g(Y) | X = x] f_X(x) dx \]

\[ E[g(X, Y)] = \int_{-\infty}^{\infty} E[g(X, Y) | X = x] f_X(x) dx \]
Continuous Bayes’ Rule

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

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

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

$$P(N = n|Y = y) = \frac{p_N(n)f_{Y|N}(y|n)}{f_Y(y)} = \frac{p_N(n)f_{Y|N}(y|n)}{\sum_i p_N(i)f_{Y|N}(y|i)}$$
Independence of Continuous RV

$X$ and $Y$ are independent if and only if

- $f_{X,Y}(x, y) = f_X(x)f_Y(y), \forall x, y$, or
- $f_{X|Y}(x|y) = f_X(x), \forall x$ and $\forall y$ such that $f_Y(y) > 0$.

If $X$ and $Y$ are independent,

- then, $g(X)$ and $h(Y)$ are independent.
- then, $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$.
- then, $\text{Var}(X + Y) = \text{Var}(X) + \text{var}(Y)$. 
Derived distributions

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

Method:

- Get the CDF:
  \[ F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = \int_{x | g(x) \leq y} f_X(x) \, dx \]

- Differentiate: \[ f_Y(y) = \frac{dF_Y}{dy}(y) \]
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
Law of iterated expectations

$E[X|Y]$ is a random variable that is a function of $Y$ (the expectation is taken with respect to $X$).

To compute $E[X|Y]$, first express $E[X|Y = y]$ as a function of $y$.

Law of iterated expectations:

$$E[X] = E[E[X|Y]]$$

(equality between two real numbers)
Law of conditional variances

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

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

as a function of $y$.

Law of conditional variances:

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

(equality between two real numbers)
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:

$$E[Y] = E[X]E[N]$$
$$\text{Var}(Y) = E[N]\text{Var}(X) + (E[X])^2\text{Var}(N)$$

Transform $M_Y(s)$ is found by starting with the transform $M_N(s)$ and replacing each occurrence of $e^s$ with $M_X(s)$. 

Bernoulli Process

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

\[
P(X_i = 1) = p \\
P(X_i = 0) = 1 - p
\]

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

  \[
p_T(t) = (1 - p)^{t-1}p, \quad t = 1, 2 \ldots
\]

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

  \[
p_K(k) = \binom{n}{k}p^k(1 - p)^{n-k}, \quad k = 0, 1 \ldots n
\]

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{align*}
  Y_k &= T_1 + T_2 \ldots T_k \\
  \mathbb{E}[Y_k] &= \mathbb{E}[T_1 + T_2 \ldots T_k] = \frac{k}{p} \\
  \text{var}(Y_k) &= \text{var}(T_1 + T_2 \ldots T_k) = \frac{k(1 - p)}{p^2}
  \end{align*}
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 independent Bernoulli processes with parameters $p$ and $q$, we get a Bernoulli process with parameter $(p+q-pq)$. 
<table>
<thead>
<tr>
<th></th>
<th>Bernoulli</th>
<th>Geometric</th>
<th>Binomial</th>
<th>Pascal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-arrival time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of arrivals</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K^{th}$ arrival</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>