**2010 Spring Quiz2 Study Notes by Joongwoo Brian Park**
Send error report to jwbpark@mit.edu

1. **Probability Density Functions (PDF)**
For a continuous RV $X$ with PDF $f_X(x)$ $(\geq 0)$,

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

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

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

2. **Mean and variance of a continuous RV**

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

3. **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:

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

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

4. **Normal/Gaussian Random Variables**
Standard Normal RV: $N(0, 1)$:

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

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

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

- 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:

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

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

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

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

$$\begin{aligned} &P(x \leq X \leq x+\delta,\ y \leq Y \leq y+\delta) \approx f_{X,Y}(x,y).\delta^2 \\ &P(A) = \int\int_A f_{X,Y}(x,y)dxdy \\ &E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dxdy \\ &f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \end{aligned}$$

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

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

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

$$\begin{aligned} f_X(x) &= \sum_{i=1}^{n} P(A_i)f_{X|A_i}(x) \ (\approx \text{total probability theorem}) \\ E[X] &= \sum_{i=1}^{n} P(A_i)E[X|A_i] \ \text{(total expectation theorem)} \\ E[g(X)] &= \sum_{i=1}^{n} P(A_i)E[g(X)|A_i] \end{aligned}$$

7. **Conditioning on a RV**
$X,\ Y$ continuous RV, $A$ an event.

$$\begin{aligned} &P(x \leq X \leq x+\delta | Y \approx y) \approx f_{X|Y}(x|y).\delta \\ &f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &f_X(x) = \int_{-\infty}^{\infty} f_Y(y)f_{X|Y}(x|y)dy \\ &P(A) = \int_{-\infty}^{\infty} P(A|X=x)f_X(x)dx \end{aligned}$$

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

8. **Continuous Bayes' Rule** $X, Y$ continuous RV, $N$ discrete RV, $A$ an event.

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)} = \frac{f_{Y|X}(y|x)f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|t)f_X(t)dt} \\ P(A|Y=y) &= \frac{P(A)f_{Y|A}(y)}{f_Y(y)} = \frac{P(A)f_{Y|A}(y)}{f_{Y|A}(y)P(A) + f_{Y|A^c}(y)P(A^c)} \end{aligned}$$

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

9. **Independence of continuous RV**

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

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

- Get the CDF:

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

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

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**

$E[X|Y]$ 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)

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

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

as a function of $y$.
Law of conditional variances:

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

(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:

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

15. **Least square prediction**

Goal: estimate RV $X$ with a real number $c$.
By definition of least square prediction, the best estimator minimizes the mean square error $E[(X-c)^2]$.
Best estimator in the absence of information: $c = E[X]$.

Corresponding mean square error: $\mathrm{Var}(X)$.

16. **Covariance and Correlation**

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

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)

17. **Uniform continuous RV over $[a,b]$**

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

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

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

18. **Exponential RV with parameter $\lambda$**

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

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

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

19. **Normal RV with parameters $(\mu, \sigma^2)$**

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

$$E[X] = \mu$$

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

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

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

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

22. **Fresh-Start** Every arrival restarts the process.

23. **Important RV associated with Bernoulli Processes**

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

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

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

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

Note: (n-fixed,k-random)

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

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

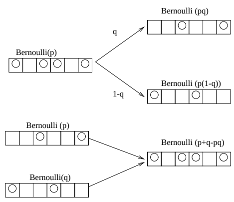
Note: (k-fixed,t-random)

24. **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

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

25. **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 an **independent** Bernoulli process with parameters pq and p(1-q)
- Conversely, if we merge two Bernoulli processes with parameters p and q, we get an **independent** Bernoulli process with parameter (p+q-pq).