Topics for Probability (the Sequel)

Disclaimer: This checklist includes topics that we have covered on probability since Midterm 3 (random variables, mean, variance, deviation bounds, sampling). It is meant to assist students in studying for the Final, but in no way is this checklist guaranteed to be comprehensive, nor is it a substitute for reviewing official course materials (e.g., lecture videos/slides, the textbook, class problems, problem sets...). The length of the sub-lists or any wording in this document should not be interpreted as suggestive regarding what problems will be on the actual Final. Content changes from version 1 (May 16) are highlighted in yellow. Content changes from version 2 (May 18) are highlighted in gray.

Random Variables

☐ Random variable

☐ variable – a variable (literally) represents how many, how long, how much.

☐ Random – the variable has a probability of taking on a specific value or range of values.

☐ Together, random variable on a probability space is a total function: domain is the sample space, codomain can be anything.

☐ In relation to events, a random variable that takes on several values partitions the sample space into blocks, with each block being an event.

☐ Functions of a random variable are themselves random variables.

☐ Indicator random variables (a.k.a. Bernoulli random variables)

☐ Maps every outcome to either 0 or 1;

☐ Two events are independent iff their indicator random variables are independent.

☐ Independent random variables: let $X$ and $Y$ be 2 random variables, $\Pr[X|Y] = \Pr[X]$ or $\Pr[Y] = 0$.

☐ Probability density function of a random variable $R$: $PDF_R(x) ::= \begin{cases} \Pr[R = x] & \text{if } x \in \text{range}(R) \\ 0 & \text{if } x \notin \text{range}(R) \end{cases}$

☐ gives the probability that $R$ takes on a particular value in its range;

☐ is always between 0 and 1;

☐ $\sum_{x \in \text{range}(R)} PDF_R(x) = 1$ for discrete $R$, $\int_{\text{smallest } x \in \text{range}(x)}^{\text{largest } x \in \text{range}(x)} PDF_R(x) = 1$ for continuous $R$

☐ Cumulative distribution function of a random variable $R$: $CDF_R(x) ::= \Pr[R \leq x]$

☐ gives the probability that $R \leq x$ (~ summing/integrating $PDF_R$ up to $x$);

☐ is always between 0 and 1, but in different way from PDF;
\[ \lim_{x \to (\infty)} CDF_R(x) = 0, \quad \lim_{x \to -\infty} CDF_R(x) = 1. \]

- **4 signature kinds of random variables**
  
  - **Uniform r.v.**
    
    - A random variable that takes on each value in its range with equal probability
    
    \[ PDF_{\text{uniform}}(x) = \begin{cases} \frac{1}{n} & \text{if } x \in \text{range}(R) \\ 0 & \text{if } x \notin \text{range}(R) \end{cases}, \] where \( n = |\text{range}(R)| \)
    
    \[ CDF_{\text{uniform}}(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{k}{n} & \text{if } k \leq x < k + 1, \text{ for } 1 \leq k < n \\ 1 & \text{if } n \leq x \end{cases} \]
  
  - **Bernoulli r.v.**
    
    - Is binary: 0 or 1, yes or no, on or off, happen or not happen
    
    - The special case “0 or 1” is also called the indicator random variable
    
    \[ PDF_{\text{Bernoulli}}(x) = \begin{cases} p & \text{if } x = 1, \text{yes, on, happen} \\ 1 - p & \text{if } x = 0, \text{no, off, not happen} \end{cases} \]
    
    \[ CDF_{\text{Indicator}}(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - p & \text{if } 0 \leq x < 1 \\ p & \text{if } 1 \leq x \end{cases} \]
  
  - **Binomial r.v.**
    
    - Such as the number of heads in \( n \) flips of a coin that comes up heads with probability \( p \)
    
    \[ PDF_{\text{binomial}}(k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k} \]
    
    \[ CDF_{\text{binomial}}(x) = \begin{cases} 0, & \text{if } x < 1 \\ \sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} \cdot p^i \cdot (1 - p)^{n-i}, & \text{if } k \leq x < k + 1, \text{ for } 1 \leq k < n \\ 1, & \text{if } n \leq x \end{cases} \]
  
  - **Time to failure r.v.**
    
    - Such as the number of coin flips up to the first head, where each coin flip is independent of other flips and comes up head with probability \( p \);
    
    - Also represents time-to-success, time-to-first-head, time-to-crash…
    
    - Is an infinite random variable! (~ it’s unlikely but not impossible to keep getting tails!)
    
    - Includes the time of the failure
    
    - Has a geometric distribution: \( \Pr[R = i] = (1 - p)^{i-1} \cdot p \)
CDF time-to-failure(t)? (~ let your summing skills for geometric series shine!)

Expectation of Random Variables

- Expectation of a random variable (a.k.a. mean, average)
  
  - Definition 1: \( E[X] := \sum_{\omega \in \text{Sample space}} R(\omega) \Pr[\omega] \)
  
  - Definition 2: \( E[X] := \sum_{x \in \text{range}(R)} x \cdot \Pr[R = x] \)

- Conditional expectation: \( E[X|A] := \sum_{x \in \text{range}(R)} x \cdot \Pr[R = x|A] \)

- Law of Total Expectation: \( E[X] = \sum_i E[X|A_i] \Pr[A_i] \), where \( R \) is a random variable in sample space \( S \) with partitions \( A_i \)

- For uniform random variable: \( E[X[R_n]] = \frac{a_1+a_2+\cdots+a_n}{n} \)
  
  - where \( R_n \) has a uniform distribution over \( \{a_1, a_2, \cdots, a_n\} \)

- For reciprocal random variable: \( E[S] = E[\frac{1}{R}] \neq \frac{1}{E[X[R]']} \), where \( S = \frac{1}{R} \)

- For indicator (a.k.a. Bernoulli) random variable: \( E[I_A] = \Pr[I_A = p] \), with \( p := \Pr[I_A = 1] \)

- For binomial random variable: \( E[I_1 + I_2 + \cdots + I_n] = \sum_{k=0}^n k \left( \begin{array}{c} n \cr k \end{array} \right) p^k (1-p)^{n-k} = np \)

- For mean time to failure random variable:
  
  - \( E[R = \# \text{ steps to failure} = \frac{1}{p} \), where \( p \) is the probability of “no failure” at individual step

- Linearity of Expectation

  - \( E[R_1 + R_2 + \cdots + R_n] = E[R_1] + E[R_2] + \cdots + E[R_n] \)

  - \( E[a_1R_1 + a_2R_2 + \cdots + a_nR_n] = a_1E[R_1] + a_2E[R_2] + \cdots + a_nE[R_n] \), where \( a_i \)'s are constant

- Applies to any kind of random variables

- Applies to sum of infinite number of random variables if \( \sum_{i=0}^\infty E[|R_i|] \) converges

- Expectation of product

  - \( E[R_1 \cdot R_2] \neq E[R_1] \cdot E[R_2], \) UNLESS \( R_1 \) and \( R_2 \) are independent or by serendipity

  - \( E[\prod_{i=1}^k R_i] \neq \prod_{i=1}^k E[R_i], \) UNLESS \( R_i \)'s are mutually independent or by serendipity

Deviation

- Variance (a.k.a. mean square deviation)
Definition: $\text{Var}[R] := \text{Ex}[(R - \text{Ex}[R])^2]$

For any random variable $R$:

- $\text{Var}[R] = \text{Ex}[R^2] - (\text{Ex}[R])^2$
- $\text{Var}[aR] = a^2\text{Var}[R]$, where $a$ is a constant (aka. Square Multiple Rule for Variance)
- $\text{Var}[R + b] = \text{Var}[R]$, where $b$ is a constant (~ constant has no variance!)

- $\text{Var}[A + B] \neq \text{Var}[A] + \text{Var}[B]$ \textbf{UNLESS} $A$ and $B$ are independent \textbf{or by serendipity}
- $\text{Var}[A + B + C + \cdots + M] \neq \text{Var}[A] + \text{Var}[B] + \text{Var}[C] + \cdots + \text{Var}[M]$ \textbf{UNLESS} $A, B, C, \ldots, M$ are pairwise independent \textbf{or by serendipity}

- $\text{Var}[AB] \neq \text{Var}[A] \cdot \text{Var}[B]$ \textbf{ALWAYS UNLESS by serendipity} (have to calculate…)

- For uniform random variable: $\text{Var}[R] = \text{Ex}[R^2] - (\text{Ex}[R])^2$ (have to calculate…)

- For reciprocal random variable: $\text{Var}[S] = \text{Var}\left[\frac{1}{R}\right]$, where $S = \frac{1}{R}$ (have to calculate…)

- For indicator (a.k.a. Bernoulli) random variable: $\text{Var}[I_A] = p(1 - p)$, with $p := \text{Pr}[I_A = 1]$

- For binomial random variable: $\text{Var}[I_1 + I_2 + \cdots + I_n] = np(1 - p)$

- For Mean Time to Failure: $\text{Var}[T] = \frac{1 - p}{p^2}$

Standard deviation $\sigma$ (a.k.a. root mean square deviation)

- Definition: $\sigma_R := \sqrt{\text{Var}[R]}$
- $\sigma_{(aR+b)} = |a|\sigma_R$

\textbf{Deviation Bounds}

- Markov Theorem
  - For nonnegative random variable $R$, $\text{Pr}[R \geq x] \leq \frac{\text{Ex}[R]}{x}$, for all $x > 0$.
  - Corollary form: $\text{Pr}[R \geq c \cdot \text{Ex}[R]] \leq \frac{1}{c}$, for all $c \geq 1$.
  - Markov bound gives a coarse estimate of the probability of a random variable taking on a value much larger than its mean.

- Chebyshev’s Theorem
  - For any random variable $R$, $\text{Pr}[(R - \text{Ex}[R]) \geq x] \leq \frac{\text{Var}[R]}{x^2}$, for all $x \in \mathbb{R}^+$. 

Page 4 of 7
Standard deviation version: \( \Pr[|R - \mathbb{E}[R]| \geq c\sigma_R] \leq \frac{1}{c^2}, \) for all \( x \in \mathbb{R}^+ \).

Chebyshev’s is based on Markov Theorem, but gives a stronger bound on the probability of a random variable deviating from its mean by at least certain amount.

**Sampling**

- Random sampling

  - How large a sample for an estimate to be within 0.04 of the real \( p \) at least 95% of time? \( 3125 \)

  - Population size does not matter!

  - Reasoning: let \( K_i \)’s be mutually independent indicator random variables, \( S_n = \sum_{i=1}^{n} K_i \)

    - \( \text{Var}[S_n] \) is maximized at \( \frac{n}{4} \) (this happens when \( p = \frac{1}{2} \); by chance is the most varied!)

    - We want: \( \Pr[\left| \frac{S_n}{n} - p \right| \leq 0.04] \geq 0.95 \) (same as \( \Pr[\left| \frac{S_n}{n} - p \right| \geq 0.04] \leq (1 - 0.95) \))

    - Apply Chebyshev’s: \( \Pr[\left| \frac{S_n}{n} - p \right| \geq 0.04] \leq \frac{\text{Var}[\frac{S_n}{n}]}{(0.04)^2} \leq \frac{1}{4n(0.04)^2} = \frac{5}{100}, \) when \( n = 3125 \)

- Pairwise independent sampling

  - Applies to any pairwise independent random variables (not necessarily 1/0 random variables) WITH the same mean and variance (but not necessarily the same distribution)

  - Let \( G_1, G_2, ..., G_n \) be pairwise independent random variables with mean \( \mu \), variance \( \sigma^2 \), and \( S_n := \sum_{i=1}^{n} G_i \),

    \[ \Pr[|\frac{S_n}{n} - \mu| \geq x] \leq \frac{1}{n} \left( \frac{\sigma^2}{x^2} \right) \] (by Chebyshev’s and Square Multiple Rule of Variance)

- Weak Law of Large Numbers

  - Meaning: by choosing a large enough sample size, we can accurately estimate the mean with confidence arbitrarily close to 100%.

  - Math: let \( G_1, G_2, ..., G_n \) be pairwise independent random variables with [same] mean \( \mu \), [finite] variance \( \sigma^2 \), and \( S_n := \sum_{i=1}^{n} G_i \)

    \[ \lim_{n \to \infty} \Pr[|\frac{S_n}{n} - \mu| \leq \epsilon] = 1 \]

- Confidence vs. Probability

  - Players: reality, our estimate of reality, our probabilistic procedure to get our estimate.

  - Confidence or confidence level says nothing about reality or how good our estimate actually is; it is about the probability that our procedure yields an estimate that is within \( \pm x \) of reality.

**Expectation for Repeated Processes**

- Framework: 2 time-to-failure random variables in play to generate a 3\textsuperscript{rd} time-to-failure random variable

- Setup: 1) flip a fair coin in independent tosses until a head (call this process \( P_0 \))

  - \( \text{where } P_0 \text{ generates a time-to-failure r.v., } T_0, \text{ the number of tails before the first head}; \)
2) repeat such process until the associated $T$ taking a value larger than the value $T_0$ took on
3) the number of repeats, $r$, is the value for the 2$^\text{nd}$ time-to-failure r.v. (call it $R$)
4) the number of total coin tosses until finishing is the 3$^\text{rd}$ time-to-failure r.v. (call it $W$)
   (no worry about $W$ for 6.042, do worry for 6.262)

- Use in 6.042?
  - to exercise time-to-failure random variables;
  - to exercise conditional expectation and the Law of Total Expectation;
  - to conclude: the expected wait time for any random variable to achieve a larger value is infinite.

- $T_0$: $\Pr[T_0 = k] = \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right) = 2^{-(k+1)}$,  $\mathbb{E}[T_0] = \left(\frac{1}{1/2}\right) - 1 = 1$

- $R$: $\Pr[R = r] = (1 - \Pr[T])^{r-1} \cdot \Pr[T] = (1 - \Pr[T_0])^{r-1} \cdot \Pr[T_0]$,  $\mathbb{E}[R | T_0 = k] = \frac{1}{\Pr[T_0 = k]}$

  By the Law of Total Expectation,
  $$\mathbb{E}[R] = \sum_{k \in \mathbb{N}} \mathbb{E}[R | T = k] \cdot \Pr[T = k] = \sum_{k \in \mathbb{N}} \mathbb{E}[R | T_0 = k] \cdot \Pr[T_0 = k] = \sum_{k \in \mathbb{N}} \frac{1}{\Pr[T_0 = k]} \cdot \Pr[T_0 = k] = \sum_{k \in \mathbb{N}} 1 = \infty$$

**Gambler’s Ruin**

- A real-world situation that illustrates concepts about 1-dimensional random walks;
- Easier to think of it as random variables lego:
  - Gambler: makes a sequence of $1$ bets  (get $1$ if wins; lose $1$ if loses)
  - Initial stake: $\$n$
  - Game endpoint: $\$T$ in hand (win) OR $\$0$ in hand (ruin)
  - Profit: $\$m$  ($m = T - n$)
  - Probability of winning each bet: $p$  ($p = \frac{1}{2}$ if game is unbiased; $p \neq \frac{1}{2}$ if game is biased)

- $\Pr[\text{Gambler wins}] = \begin{cases} \frac{n}{T} & \text{for } p = \frac{1}{2}, \\
\frac{r^{n-1}}{r^{T-1}} & \text{for } p \neq \frac{1}{2}, \end{cases}$  
  (this is a uniform distribution)

  (derived from Pascal’s re-assignment of bet-wise gain/loss to simulate a fair game)

  where $r := \frac{q}{p}$ and $q := (1 - p)$

- Probability of winning reasoned from recurrence
  - Treat each bet as a “new beginning” of the game, with the money in hand as the new initial stake;
  - Note, this “new beginning” will, based on $p$ and $q$, give rise to the next “new beginning”
  - Math, let $w_n$ be the gambler’s probability of winning when his initial stake is $\$n$:
    - initial conditions: $w_0 = 0$,  $w_T = 1$
    - recurrence: $w_n = p \cdot w_{n+1} + q \cdot w_{n-1}$, so $w_{n+1} = \frac{1}{p} \cdot w_n - \frac{q}{p} \cdot w_{n-1}$
Bound on the probability of winning in the biased game with \( p < \frac{1}{2} \): \( \Pr[\text{Gambler wins}] < \left( \frac{p}{q} \right)^{T-n} \)

Observations:
- \( p \uparrow \), bound on probability of winning \( \uparrow \);
- \( (T - n) \uparrow \), bound on probability of winning \( \uparrow \), independent of specific values of \( T \) or \( n \) (only their difference, the intended profit, matters)!

The probability of winning in the game even just slightly biased against the gambler is SMALL:

- Let \( k \leq \min(m, n) \). After \( k \) bets, the number of wins by the gambler has a binomial distribution with \( p < \frac{1}{2} \).

For the gambler to win, he needs the number of winning bets to deviate by

\[
\frac{m + k(1 - 2p)}{\sqrt{kp(1 - 2p)}} = \Theta(\sqrt{k})
\]

times its \( \sigma \), from its mean: \( m + k(1 - 2p) \). This number is very unlikely to reach.

\[
E_x[\text{number of bets}] = \begin{cases} 
    n(T - n) & \text{for } p = \frac{1}{2} \\
    \frac{w_n(T - n)}{p - q} & \text{for } p \neq \frac{1}{2}, \text{ where } w_n = \Pr[\text{Gambler wins}] = (r^n - 1)/(r^T - 1)
\end{cases}
\]

Keep playing, gambler’s ruin is guaranteed!

If the gambler starts with \( n \geq 1 \) and play a fair unbounded game, then

- he will go broke with probability 1;
- his expected number of plays is infinite.