CMPT 210: Probability and Computation

Lecture 24

Sharan Vaswani

August 5, 2022

Logistics

Final Exam is on August 14 (Sunday) from 12 pm - 3 pm in AQ 3005.

Scope of the Final:

- Syllabus includes everything that we have covered (Lectures 1 - 24 and Assignments 1-4).
- For continuous r.v's, there will be only very basic questions (no difficult integrals).

You are allowed to bring a A4-sized, two-sided, hand-written formula sheet for the Final.

Go through the slides/assignments and (Meyer, Lehman, Leighton) to prepare.

Final will be "easy" – if your concepts are clear, you should be able to get full marks.

Office hours next week: Tuesday, 9 August, 11 am - 1 pm & Thursday, 11 August, 9 am - 10 am.

Recap

The distribution of a continuous r.v. R is completely specified by its PDF $f_R : \mathbb{R} \to \mathbb{R}_+$ and CDF $F_R : \mathbb{R} \to [0,1]$.

Probability Density Function: For all u, $f_R(u) \geq 0$ and satisfies $\Pr[R \in [a,b]] = \int_a^b f_R(u)\, du$. $\int_{-\infty}^{\infty} f_R(u)\, du = 1$.

Cumulative Distribution Function: For all u, $F_R(u) := \Pr[R \leq u] = \int_{-\infty}^{u} f_R(u)\, du$ and satisfies: $\lim_{u \to -\infty} F_R(u) = 0$ and $\lim_{u \to \infty} F_R(u) = 1$.

PDF and CDF: For any continuous r.v. R, $\frac{dF_R(v)}{dv} = \frac{d \int_{-\infty}^{v} f_R(u)\, du}{dv} = f_R(v)$.

Expectation and Variance: For a continuous r.v. R, $\mathbb{E}[R] = \int_{-\infty}^{\infty} u\, f_R(u)\, du$ and $\mathrm{Var}[R] = (\int_{-\infty}^{\infty} u^2\, f_R(u)\, du) - (\int_{-\infty}^{\infty} u\, f_R(u)\, du)^2$.

Recap

Continuous uniform distribution: If $R \sim \text{Uniform}[a, b]$, for all $u \in [a, b]$, $f_R(u) = \frac{1}{b-a}$ and $f_R(u) = 0$ if $u \notin [a, b]$. $\forall u \in [a, b]$, $F_R(u) = \frac{u-a}{b-a}$. $F_R(u) = 0$ if $u < a$ and $F_R(u) = 1$ if $u > b$.

Expectation and Variance for the continuous uniform distribution: If $R \sim \text{Uniform}[a, b]$, $\mathbb{E}[R] = \frac{b+a}{2}$ and $\text{Var}[R] = \frac{(b-a)^2}{12}$.

Standard Normal Distribution: Random variable R follows the standard normal distribution i.e. $X \sim \mathcal{N}(0, 1)$ if $f_R(u) = \Phi(u) := \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-u^2}{2}\right)$.

Normal Distribution: Random variable R follows the Normal distribution i.e. $R \sim \mathcal{N}(\mu, \sigma^2)$ if $f_R(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$.

Expectation and Variance for the normal distribution: If $R \sim \mathcal{N}(\mu, \sigma^2)$, $\mathbb{E}[R] = \mu$ and $\text{Var}[R] = \sigma^2$.

Standardizing a Gaussian: If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$.

Questions?

Properties of the Normal Distribution

Sum of independent Gaussian r.v's: If $X_1, X_2, \ldots, X_n$ are mutually independent random variables, and $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, then if $X = X_1 + X_2 + \ldots + X_n$, then $X \sim \mathcal{N}\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$.

As a check, note that by the linearity of expectation, $\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n \mu_i$.

Similarly, by the linearity of variance of pairwise independent random variables, $\text{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \text{Var}[X_i] = \sum_{i=1}^n \sigma_i^2$.

The above statement is much stronger – not only does it quantify the mean and variance of the sum of independent Gaussian r.v's, it also says that the resulting distribution of X is also a Gaussian!

Central Limit Theorem

We have seen that the normal distribution can be seen as the limit of the Binomial distribution – specifically, for large n, if $X_1, X_2, \ldots, X_n$ are Bernoulli random variables with parameter p, then for $X = X_1 + X_2 + \ldots X_n$, $f_X(x) \approx \sqrt{\frac{1}{2\pi\,\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ where $\mu = \mathbb{E}[X] = np$ and $\sigma^2 = \mathrm{Var}[X] = n\,p\,(1-p)$.

We also saw that if $X_1, X_2, \ldots, X_n$ are independent Gaussian r.v's (with mean μ_i and variance σ_i^2) and $X = X_1 + X_2 + \ldots + X_n$, then, $f_X(x) = \sqrt{\frac{1}{2\pi\,\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ where $\mu = \sum_i \mu_i$ and $\sigma^2 = \sum_i \sigma_i^2$.

Hence, in both cases, by "standardizing" X i.e. for $Y := \frac{X-\mu}{\sigma}$, $Y \sim \mathcal{N}(0,1)$.

Central Limit Theorem

Central Limit Theorem: For independent random variables $X_1, X_2, \ldots, X_n$ with finite mean $\mu := \mathbb{E}[X_i]$ and finite variance $\sigma^2 := \text{Var}[X_i]$, if $X = X_1 + X_2 + \ldots + X_n$ and $Y := \frac{X - n\mu}{\sqrt{n}\sigma}$ (such that $\mathbb{E}[Y] = 1$ and $\text{Var}[Y] = 1$), then, for all t,

$$\lim_{n\to\infty} F_Y(t) = \lim_{n\to\infty} \Pr[Y \le t] = \Phi(t) = \Pr[\mathcal{N}(0,1) \le t] = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$$

This is true for **any** distribution of the X_i's! (given that the mean and variances are bounded), but is only an asymptotic result (only true as $n \to \infty$).

Compare this to the Chernoff bound that is non-asymptotic (holds for all n and has an explicit dependence on n) (and requires the $X_i \in [0,1]$). Chernoff only bounds the probability of deviation from the mean, while CLT is a statement about the whole distribution.

Compare this to the weak law of large numbers that proves that $\lim_{n\to\infty} X/n = \mu$ and is an asymptotic statement about the mean. On the other hand, CLT is a statement about the whole distribution.

Central Limit Theorem

In practice, for large n (when $n \gtrapprox 30$), the CLT is a powerful tool – by bounding the CDF of a Gaussian, we can obtain a handle on the distribution of Y and hence X. It can thus be used as an alternate to the tail inequalities we discussed earlier.

Under additional assumptions, CLT can be modified to give a non-asymptotic bound in the form of the Berry-Esseen Theorem.

Berry-Esseen Theorem: For independent random variables $X_1, X_2, \ldots, X_n$ with finite mean $\mu := \mathbb{E}[X_i]$ and finite variance $\sigma^2 := \text{Var}[X_i]$, if $X = X_1 + X_2 + \ldots + X_n$ and $Y := \frac{X - n\mu}{\sqrt{n}\sigma}$ (such that $\mathbb{E}[Y] = 1$ and $\text{Var}[Y] = 1$) and $\beta := \mathbb{E}[|X]|^3] < \infty$, then, for all t,

$$|F_Y(t) - \Phi(t)| \leq O\left(\frac{\beta}{\sqrt{n}}\right).$$

Hence, under the additional assumption that the third moment is bounded, the distribution of Y approaches that of the standard normal distribution at an $O(1/\sqrt{n})$ rate.

The Berry-Esseen theorem gives some justification why the CLT works so well for the well-behaved real distributions even for finite n.

Questions?

Wrapping up

Sample (outcome) space $\mathcal{S}$: Nonempty (countable) set of possible outcomes.

Outcome $\omega \in \mathcal{S}$: Possible "thing" that can happen.

Event E: Any subset of the sample space.

Probability function on a sample space $\mathcal{S}$ is a total function $\Pr : \mathcal{S} \to [0, 1]$. For any $\omega \in \mathcal{S}$,

$$0 \leq Pr[\omega] \leq 1 \quad ; \quad \sum_{\omega \in \mathcal{S}} \Pr[\omega] = 1 \quad ; \quad \Pr[E] = \sum_{\omega \in E} \Pr[\omega]$$

Wrapping up

Union: For mutually exclusive events $E_1, E_2, \ldots, E_n$, $\Pr[E_1 \cup E_2 \cup \ldots E_n] = \Pr[E_1] + \Pr[E_2] + \ldots + \Pr[E_n]$.

Complement rule: $\Pr[E] = 1 - \Pr[E^c]$

Inclusion-Exclusion rule: For any two events E, F, $\Pr[E \cup F] = \Pr[E] + \Pr[F] - \Pr[E \cap F]$.

Union Bound: For any events $E_1, E_2, E_3, \ldots E_n$, $\Pr[E_1 \cup E_2 \cup E_3 \ldots \cup E_n] \leq \sum_{i=1}^{n} \Pr[E_i]$.

Uniform probability space: A probability space is said to be uniform if $\Pr[\omega]$ is the same for every outcome $\omega \in \mathcal{S}$. In this case, $\Pr[E] = \frac{|E|}{|\mathcal{S}|}$.

Wrapping up

Conditional Probability: For events E and F, probability of event E conditioned on F is given by $\Pr[E|F]$ and can be computed as $\Pr[E|F] = \frac{\Pr[E \cap F]}{\Pr[F]}$.

Probability rules with conditioning: For the complement E^c, $\Pr[E^c|F] = 1 - \Pr[E|F]$.

Conditional Probability for multiple events:
$\Pr[E_1 \cap E_2 \cap E_3] = \Pr[E_1] \Pr[E_2|E_1] \Pr[E_3|E_1 \cap E_2]$.

Bayes rule: For events E and F if $\Pr[E] \neq 0$, $\Pr[F|E] = \frac{\Pr[E|F] \Pr[F]}{\Pr[E]}$.

Law of Total Probability: For events E and F, $\Pr[E] = \Pr[E|F] \Pr[F] + \Pr[E|F^c] \Pr[F^c]$.

Independent Events: Events E and F are said to be independent, if knowledge that F has occurred does not change the probability that E occurs, i.e. $\Pr[E|F] = \Pr[E]$ and $\Pr[E \cap F] = \Pr[E] \Pr[F]$.

Wrapping up

Pairwise Independence: Events E_1, E_2, ..., E_n are pairwise independent, if for *every* pair of events E_i and E_j ($i \neq j$), $\Pr[E_i | E_j] = \Pr[E_i]$ and $\Pr[E_i \cap E_j] = \Pr[E_i] \Pr[E_j]$.

Mutual Independence: Events E_1, E_2, ..., E_n are mutually independent, if for *every* subset of events, the probability that all the selected events occur equals the product of the probabilities of the selected events. Formally, for every subset $S \subseteq \{1, 2, \ldots, n\}$, $\Pr[\cap_{i \in S} E_i] = \prod_{i \in S} \Pr[E_i]$.

Random variable: A random "variable" R on a probability space is a total function whose domain is the sample space $\mathcal{S}$. The codomain is denoted by V (usually a subset of the real numbers), meaning that $R : \mathcal{S} \rightarrow V$.

Indicator Random Variables: An indicator random variable corresponding to an event E is denoted as $\mathcal{I}_E$ and is defined such that for $\omega \in E$, $\mathcal{I}_E[\omega] = 1$ and for $\omega \notin E$, $\mathcal{I}_E[\omega] = 0$.

Wrapping up

Probability density function (PDF): Let R be a random variable with codomain V. The probability density function of R is the function $\text{PDF}_R : V \to [0, 1]$, such that $\text{PDF}_R[x] = \Pr[R = x]$ if $x \in \text{Range(R)}$ and equal to zero if $x \notin \text{Range(R)}$.

$\sum_{x \in V} \text{PDF}_R[x] = \sum_{x \in \text{Range(R)}} \Pr[R = x] = 1$.

Cumulative distribution function (CDF): The cumulative distribution function of R is the function $\text{CDF}_R : \mathbb{R} \to [0, 1]$, such that $\text{CDF}_R[x] = \Pr[R \leq x]$.

Distribution over a random variable can be fully specified using the cumulative distribution function (CDF) (usually denoted by F). The corresponding probability density function (PDF) is denoted by f.

Wrapping up

Bernoulli Distribution: $f_p(0) = 1 - p$, $f_p(1) = p$. *Example*: When tossing a coin such that $\Pr[\text{heads}] = p$, random variable R is equal to 1 if we get a heads (and equal to 0 otherwise). In this case, R follows the Bernoulli distribution i.e. $R \sim \text{Ber}(p)$.

Uniform Distribution: If $R : \mathcal{S} \to V$, then for all $v \in V$, $f(v) = 1/|V|$. *Example*: When throwing an n-sided die, random variable R is the number that comes up on the die. $V = \{1, 2, \ldots, n\}$. In this case, R follows the Uniform distribution i.e. $R \sim \text{Uniform}\{1, 2, \ldots, n\}$.

Binomial Distribution: $f_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}$. *Example*: When tossing n independent coins such that $\Pr[\text{heads}] = p$, random variable R is the number of heads in n coin tosses. In this case, R follows the Binomial distribution i.e. $R \sim \text{Bin}(n, p)$.

Geometric Distribution: $f_p(k) = (1-p)^{k-1} p$. *Example*: When repeatedly tossing a coin such that $\Pr[\text{heads}] = p$, random variable R is the number of tosses needed to get the first heads. In this case, R follows the Geometric distribution i.e. $R \sim \text{Geo}(p)$.

Wrapping up

Expectation/mean of a random variable R is denoted by $\mathbb{E}[R]$ and "summarizes" its distribution. Formally, $\mathbb{E}[R] := \sum_{\omega \in \mathcal{S}} \Pr[\omega]\, R[\omega]$

Alternate definition of expectation: $\mathbb{E}[R] = \sum_{x \in \text{Range}(R)} x \, \Pr[R = x]$.

Expectation of transformed r.v's: For a random variable $X : \mathcal{S} \to V$ and a function $g : V \to \mathbb{R}$, we define $\mathbb{E}[g(X)]$ as follows: $\mathbb{E}[g(X)] := \sum_{x \in \text{Range}(X)} g(x) \Pr[X = x]$

Linearity of Expectation: For n random variables $R_1, R_2, \ldots, R_n$ and constants $a_1, a_2, \ldots, a_n$, $b_1, b_2, \ldots, b_n$, $\mathbb{E}\left[\sum_{i=1}^{n} a_i R_i + b_i\right] = \sum_{i=1}^{n} a_i \, \mathbb{E}[R_i] + b_i$.

Conditional Expectation: For random variable R, the expected value of R conditioned on an event A is given by $\mathbb{E}[R|A] = \sum_{x \in \text{Range}(R)} x \, \Pr[R = x|A]$

Law of Total Expectation: If R is a random variable $\mathcal{S} \to V$ and events $A_1, A_2, \ldots A_n$ form a partition of the sample space, then, $\mathbb{E}[R] = \sum_i \mathbb{E}[R|A_i] \, \Pr[A_i]$.

Wrapping up

Independent random variables: We define two random variables R_1 and R_2 to be independent if for *all* $x_1 \in \text{Range}(R_1)$ and $x_2 \in \text{Range}(R_2)$, events $[R_1 = x_1]$ and $[R_2 = x_2]$ are independent. More formally,

$$\Pr[(R_1 = x_1) \cap (R_2 = x_2)] = \Pr[(R_1 = x_1)] \Pr[(R_2 = x_2)]$$

Independent random variables: Two random variables R_1 and R_2 are independent if for *all* $x_1 \in \text{Range}(R_1)$ and $x_2 \in \text{Range}(R_2)$,

$$\Pr[(R_1 = x_1) | (R_2 = x_2)] = \Pr[(R_1 = x_1)]$$
$$\Pr[(R_2 = x_2) | (R_1 = x_1)] = \Pr[(R_2 = x_2)]$$

Expectation of product of r.v's: For two r.v's R_1 and R_2, $\mathbb{E}[R_1\, R_2] = \sum_{x \in \text{Range}(R_1 R_2)} x \Pr[R_1 R_2 = x]$.

Expectation of product of independent r.v's: For independent r.v's R_1 and R_2, $\mathbb{E}[R_1\, R_2] = \mathbb{E}[R_1]\, \mathbb{E}[R_2]$.

Wrapping up

Joint distribution between r.v's X and Y can be specified by its joint PDF as follows: $\text{PDF}_{X,Y}[x,y] = \Pr[X = x \cap Y = y]$.

If X and Y are independent random variables, $\text{PDF}_{X,Y}[x,y] = \text{PDF}_X[x]\,\text{PDF}_Y[y]$.

Marginalization: We can obtain the distribution for each r.v. from the joint distribution by marginalizing over the other r.v's i.e. $\text{PDF}_X[x] = \sum_i \text{PDF}_{X,Y}[x, y_i]$.

Variance: Standard way to measure the deviation from the mean. For r.v. X, $\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{x \in \text{Range}(X)} (x - \mu)^2 \Pr[X = x]$ where $\mu := \mathbb{E}[X]$.

Alternate definition of variance: $\text{Var}[X] = \mathbb{E}[X^2] - \mu^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Standard Deviation: For r.v. X, the standard deviation of X is defined as $\sigma_X := \sqrt{\text{Var}[X]} = \sqrt{\mathbb{E}[X^2] - (\mathbb{E}[X])^2}$.

Wrapping up

Properties of variance: For constants a, b and r.v. R, $\text{Var}[aR + b] = a^2\text{Var}[R]$.

Pairwise Independence of r.v's: Random variables $R_1, R_2, R_3, \ldots R_n$ are pairwise independent if for any pair R_i and R_j, for $x \in \text{Range}(R_i)$ and $y \in \text{Range}(R_j)$, $\Pr[(R_i = x) \cap (R_j = y)] = \Pr[R_i = x] \Pr[R_j = y]$.

Linearity of variance for pairwise independent r.v's: If $R_1, \ldots, R_n$ are pairwise independent, $\text{Var}[R_1 + R_2 + \ldots R_n] = \sum_{i=1}^{n} \text{Var}[R_i]$.

Properties of variance: If $R_1, \ldots, R_n$ are pairwise independent, for constants $a_1, a_2, \ldots a_n$ and $b_1, b_2, \ldots b_n$, $\text{Var}[\sum_{i=1}^{n} a_i R_i + b_i] = \sum_{i=1}^{n} a_i^2 \text{Var}[R_i]$.

Wrapping up

Covariance: For two random variables R and S, the covariance between R and S is defined as: $\mathrm{Cov}[R, S] = \mathbb{E}[(R - \mathbb{E}[R])\,(S - \mathbb{E}[S])] = \mathbb{E}[RS] - \mathbb{E}[R]\,\mathbb{E}[S]$.

Properties of covariance: If R and S are independent r.v's, $\mathbb{E}[RS] = \mathbb{E}[R]\,\mathbb{E}[S]$ and $\mathrm{Cov}[R, S] = 0$. $\mathrm{Cov}[R, R] = \mathrm{Var}[R]$. $\mathrm{Cov}[R, S] = \mathrm{Cov}[S, R]$.

Variance of sum of r.v's: For r.v's $R_1, R_2, \ldots, R_n$, $\mathrm{Var}\left[\sum_{i=1}^{n} R_i\right] = \sum_{i=1}^{n} \mathrm{Var}[R_i] + 2\sum_{1 \leq i < j \leq n} \mathrm{Cov}[R_i, R_j]$.

If R_i and R_j are pairwise independent, $\mathrm{Cov}[R_i, R_j] = 0$ and $\mathrm{Var}\left[\sum_{i=1}^{n} R_i\right] = \sum_{i=1}^{n} \mathrm{Var}[R_i]$.

Correlation: For two r.v's R_1 and R_2, the correlation between R_1 and R_2 is defined as $\mathrm{Corr}[R_1, R_2] = \frac{\mathrm{Cov}[R_1, R_2]}{\sqrt{\mathrm{Var}[R_1]\,\mathrm{Var}[R_2]}}$. $\mathrm{Corr}[R_1, R_2] \in [-1, 1]$ and indicates the strength of the relationship between R_1 and R_2.

Wrapping up

Bernoulli: If $R \sim \text{Bernoulli}(p)$, $\mathbb{E}[R] = p$ and $\text{Var}[R] = p\,(1-p)$.

Uniform: If $R \sim \text{Uniform}(\{v_1, \ldots, v_n\})$, $\mathbb{E}[R] = \frac{v_1 + v_2 + \ldots + v_n}{n}$ and $\text{Var}[R] = \frac{[v_1^2 + v_2^2 + \ldots v_n^2]}{n} - \left(\frac{[v_1 + v_2 + \ldots v_n]}{n}\right)^2$.

Binomial: If $R \sim \text{Bin}(n, p)$, $\mathbb{E}[R] = np$ and $\text{Var}[R] = n\,p\,(1-p)$.

Geometric: If $R \sim \text{Geo}(p)$, $\mathbb{E}[R] = \frac{1}{p}$ and $\text{Var}[R] = \frac{1-p}{p^2}$.

Wrapping up

Tail inequalities bound the probability that the r.v. takes a value much different from its mean.

Markov's Theorem: If X is a non-negative random variable, then for all $x > 0$, $\Pr[X \geq x] \leq \frac{\mathbb{E}[X]}{x}$.

Chebyshev's Theorem: For a r.v. X and all $x > 0$, $\Pr[|X - \mathbb{E}[X]| \geq x] \leq \frac{\mathrm{Var}[X]}{x^2}$.

Weak Law of Large Numbers: Let $G_1, G_2, \ldots, G_n$ be pairwise independent variables with the same mean μ and (finite) standard deviation σ. Define $T_n := \frac{\sum_{i=1}^n G_i}{n}$, then for every $\epsilon > 0$, $\lim_{n \to \infty} \Pr[|T_n - \mu| \leq \epsilon] = 1$.

Chernoff Bound: If $T_1, T_2, \ldots, T_n$ are mutually independent r.v's such that $0 \leq T_i \leq 1$ for all i. If $T := \sum_{i=1}^n T_i$, for all $c \geq 1$ and $\beta(c) := c \ln(c) - c + 1$, $\Pr[T \geq c\mathbb{E}[T]] \leq \exp(-\beta(c)\,\mathbb{E}[T])$.

Two-sided Chernoff Bound: $\Pr\left[|T - \mathbb{E}[T]| \geq c\mathbb{E}[T]\right] \leq 2\exp\left(\frac{-c^2\,\mathbb{E}[T]}{3}\right)$

Wrapping up

The distribution of a continuous r.v. R is completely specified by its PDF $f_R : \mathbb{R} \to \mathbb{R}_+$ and CDF $F_R : \mathbb{R} \to [0, 1]$.

Probability Density Function: For all u, $f_R(u) \geq 0$ and satisfies $\Pr[R \in [a, b]] = \int_a^b f_R(u)\, du$. $\int_{-\infty}^{\infty} f_R(u)\, du = 1$.

Cumulative Distribution Function: For all u, $F_R(u) := \Pr[R \leq u] = \int_{-\infty}^{u} f_R(u)\, du$ and satisfies: $\lim_{u \to -\infty} F_R(u) = 0$ and $\lim_{u \to \infty} F_R(u) = 1$.

PDF and CDF: For any continuous r.v. R, $\frac{dF_R(v)}{dv} = \frac{d \int_{-\infty}^{v} f_R(u)\, du}{dv} = f_R(v)$.

Expectation and Variance: For a continuous r.v. R, $\mathbb{E}[R] = \int_{-\infty}^{\infty} u\, f_R(u)\, du$ and $\text{Var}[R] = (\int_{-\infty}^{\infty} u^2\, f_R(u)\, du) - (\int_{-\infty}^{\infty} u\, f_R(u)\, du)^2$.

Wrapping up

Continuous uniform distribution: If $R \sim \text{Uniform}[a,b]$, for all $u \in [a,b]$, $f_R(u) = \frac{1}{b-a}$ and $f_R(u) = 0$ if $u \notin [a,b]$. $\forall u \in [a,b]$, $F_R(u) = \frac{u-a}{b-a}$. $F_R(u) = 0$ if $u < a$ and $F_R(u) = 1$ if $u > b$.

Expectation and Variance for the continuous uniform distribution: If $R \sim \text{Uniform}[a,b]$, $\mathbb{E}[R] = \frac{b+a}{2}$ and $\text{Var}[R] = \frac{(b-a)^2}{12}$.

Standard Normal Distribution: Random variable R follows the standard normal distribution i.e. $X \sim \mathcal{N}(0,1)$ if $f_R(u) = \Phi(u) := \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-u^2}{2}\right)$.

Normal Distribution: Random variable R follows the Normal distribution i.e. $R \sim \mathcal{N}(\mu, \sigma^2)$ if $f_R(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$.

Expectation and Variance for the normal distribution: If $R \sim \mathcal{N}(\mu, \sigma^2)$, $\mathbb{E}[R] = \mu$ and $\text{Var}[R] = \sigma^2$.

Standardizing a Gaussian: If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.

Sum of independent Gaussian r.v's: If $X_1, X_2, \ldots, X_n$ are mutually independent random variables, and $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, then if $X = \sum_{i=1}^n X_i$, then $X \sim \mathcal{N}\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$.

Wrapping up

Central Limit Theorem: For independent random variables $X_1, X_2, \ldots, X_n$ with finite mean $\mu := \mathbb{E}[X_i]$ and finite variance $\sigma^2 := \text{Var}[X_i]$, if $X = X_1 + X_2 + \ldots + X_n$ and $Y := \frac{X - n\mu}{\sqrt{n}\sigma}$ (such that $\mathbb{E}[Y] = 1$ and $\text{Var}[Y] = 1$), then, for all t,
$\lim_{n\to\infty} F_Y(t) = \lim_{n\to\infty} \Pr[Y \leq t] = \Phi(t) = \Pr[\mathcal{N}(0,1) \leq t] = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du.$

CLT holds for **any** distribution of the X_i's. (given that the mean and variances are bounded), but is only an asymptotic result (only true as $n \to \infty$).

In practice, for large n (when $n \gtrapprox 30$), the CLT is a powerful tool – by bounding the CDF of a Gaussian, we can obtain a handle on the distribution of Y and hence X. It can thus be used as an alternate to the tail inequalities we discussed earlier.

Berry-Esseen Theorem: For independent random variables $X_1, X_2, \ldots, X_n$ with finite mean $\mu := \mathbb{E}[X_i]$ and finite variance $\sigma^2 := \text{Var}[X_i]$, if $X = X_1 + X_2 + \ldots + X_n$ and $Y := \frac{X - n\mu}{\sqrt{n}\sigma}$ (such that $\mathbb{E}[Y] = 1$ and $\text{Var}[Y] = 1$) and $\beta := \mathbb{E}[|X]|^3] < \infty$, then, for all t,
$|F_Y(t) - \Phi(t)| \leq O\left(\frac{\beta}{\sqrt{n}}\right).$

What's Next?

STAT 271: Probability and Statistics for Computing Science (Offered in Fall'22)

- More continuous distributions and random variables
- Sampling and Parameter estimation
- Linear Regression
- Hypothesis testing
- Analysis of Variance

Questions?