8 Sums of Random Variables

- Consider n random variables $X_1, X_2, \dots, X_n$.
- We are often interested in the behavior of their sum $S_n = \sum_{i=1}^{n} X_i$.
- For example, the **sample mean** $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is used to estimate the mean from data.
- Unfortunately, computing the distribution of the sum S_n or sample mean M_n directly can be quite challenging. As a starting point, we can compute the expected value and variance.
- **Expected Value of the Sum:** $\mathbb{E}\big[S_n\big] = \mathbb{E}\Big[\sum_{i=1}^{n} X_i\Big] = \sum_{i=1}^{n} \mathbb{E}[X_i]$
- **Variance of the Sum:** $\mathsf{Var}\big[S_n\big] = \mathsf{Var}\Big[\sum_{i=1}^{n} X_i\Big] = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathsf{Cov}[X_i, X_j]$

8.1 Independent and Identically Distributed Random Variables

- Random variables $X_1, \dots, X_n$ are said to be **independent and identically distributed (i.i.d.)** if they are independent and all X_i have the same marginal distribution, which is a PMF $P_X(x)$ in the discrete case and a PDF $f_X(x)$ in the continuous case. The joint PMF or joint PDF can be factored as follows:
  - Discrete: $P_{X_1,\dots,X_n}(x_1, \dots, x_n) = P_X(x_1) \cdots P_X(x_n) = \prod_{i=1}^{n} P_X(x_i)$
  - Continuous: $f_{X_1,\dots,X_n}(x_1, \dots, x_n) = f_X(x_1) \cdots f_X(x_n) = \prod_{i=1}^{n} f_X(x_i)$
- For i.i.d. $X_1, \dots, X_n$, there are simple formulas for the mean and variance of sums:

$$\mathbb{E}\big[S_n\big] = \mathbb{E}\Big[\sum_{i=1}^{n} X_i\Big] = n\mathbb{E}[X] \qquad \mathsf{Var}\big[S_n\big] = \mathsf{Var}\Big[\sum_{i=1}^{n} X_i\Big] = n\mathsf{Var}[X]$$

$$\mathbb{E}\big[M_n\big] = \mathbb{E}\Big[\frac{1}{n}\sum_{i=1}^{n} X_i\Big] = \mathbb{E}[X] \qquad \mathsf{Var}\big[M_n\big] = \mathsf{Var}\Big[\frac{1}{n}\sum_{i=1}^{n} X_i\Big] = \frac{1}{n}\mathsf{Var}[X]$$

8.2 Laws of Large Numbers

- Formally, an **infinite sequence of random variables** $X_1, X_2, \dots$ is specified by a collection of joint CDFs $F_{X_{k_1},\dots,X_{k_n}}(x_{k_1}, \dots, x_{k_n})$ for any finite set of distinct indices $k_1, \dots, k_n$ and any finite n. In the discrete case, each joint CDF corresponds to a joint PMF $P_{X_{k_1},\dots,X_{k_n}}(x_{k_1}, \dots, x_{k_n})$, and, in the continuous case, to a joint PDF $f_{X_{k_1},\dots,X_{k_n}}(x_{k_1}, \dots, x_{k_n})$. If the sequence is i.i.d., then the joint distributions factor
  - Discrete: $P_{X_{k_1},\dots,X_{k_n}}(x_{k_1}, \dots, x_{k_n}) = P_X(x_{k_1}) \cdots P_X(x_{k_n})$ with marginal PMF $P_X(x)$
  - Continuous: $f_{X_{k_1},\dots,X_{k_n}}(x_{k_1}, \dots, x_{k_n}) = f_X(x_{k_1}) \cdots f_X(x_{k_n})$ with marginal PDF $f_X(x)$

- Intuition: For an i.i.d. sequence of random variables, the sample mean converges to the true mean as $n \to \infty$. Since the variance of the sample mean decreases with n, we expect our estimate of the mean to become increasingly accurate as n increases. The laws of large numbers make this intuition precise.

- **Weak Law of Large Numbers:** Let $M_n = \frac{1}{n}\sum_{i=1}^n X_i$ be the sample mean of an i.i.d. sequence of random variables $X_1, X_2, \dots$ with finite mean $\mathbb{E}[X_i] = \mu < \infty$. For any $\epsilon > 0$,
$$\lim_{n\to\infty} \mathbb{P}\big[|M_n - \mu| > \epsilon\big] = 0 \ .$$
Equivalently,
$$\lim_{n\to\infty} \mathbb{P}\big[\mu - \epsilon \leq M_n \leq \mu + \epsilon\big] = 1 \ .$$
  - Intuition: For any tolerance $\epsilon > 0$, the sample mean M_n eventually lands in the interval $[\mu - \epsilon, \mu + \epsilon]$ where μ is the true mean.
  - We can characterize how quickly the sample mean converges by imposing additional assumptions.
    - **Chebyshev's Inequality:** If $\mathsf{Var}[X_i] = \sigma^2$ is finite, then
$$\mathbb{P}\big[|M_n - \mu| > \epsilon\big] \leq \frac{\mathsf{Var}[X]}{n} \ .$$
    - **Hoeffding's Inequality:** If the random variables are bounded $a \leq X_i \leq b$, then
$$\mathbb{P}\big[|M_n - \mu| > \epsilon\big] \leq 2\exp\left(-\frac{2n\epsilon^2}{(b-a)^2}\right) \ .$$
    - **Gaussian Tail Bound:** If the X_i are Gaussian(μ, σ^2), then
$$\mathbb{P}\big[|M_n - \mu| > \epsilon\big] \leq 2\exp\left(-\frac{n\epsilon^2}{2\sigma^2}\right) \ .$$

- **Strong Law of Large Numbers:** Let $M_n = \frac{1}{n}\sum_{i=1}^n X_i$ be the sample mean of an i.i.d. sequence of random variables $X_1, X_2, \dots$ with finite mean $\mathbb{E}[X_i] = \mu < \infty$. Then,
$$\mathbb{P}\big[\lim_{n\to\infty} M_n = \mu\big] = 1 \ .$$
  - Intuition: The sample mean M_n eventually converges exactly to the true mean μ.

8.3 Central Limit Theorem

- **Central Limit Theorem:** Let $M_n = \frac{1}{n}\sum_{i=1}^n X_i$ be the sample mean of an i.i.d. sequence of random variables $X_1, X_2, \dots$ with finite mean $\mu < \infty$ and finite variance $\sigma^2 < \infty$. Then, for any value y, the CDF of $Y_n = \dfrac{\sqrt{n}(M_n - \mu)}{\sigma}$ converges to the standard normal CDF, $\lim_{n\to\infty} F_{Y_n}(y) = \Phi(y)$.

- Intuition: $\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$ looks like a Gaussian random variable for large n. That is, the sum of many small, independent effects looks Gaussian. The normalization by $\frac{1}{\sqrt{n}}$ is important to obtain convergence to a Gaussian distribution: if we instead normalize by $\frac{1}{n}$, then we get back to $\frac{1}{n}\sum_{i=1}^n X_i$, which converges to the mean μ.

- We often use the Central Limit Theorem as justification for approximating distributions by a Gaussian. Specifically, if $n > 30$, then $F_{Y_n}(y) \approx \Phi(y)$ is a good approximation. Equivalently, $F_{M_n}(m) = \mathbb{P}[M_n \leq m] \approx \Phi\left(\dfrac{m - \mu}{\sqrt{\sigma^2/n}}\right)$ and $\mathbb{P}\big[|M_n - \mu| > \epsilon\big] \approx 2Q\left(\dfrac{\epsilon\sqrt{n}}{\sigma}\right)$.