Plan for the next 8 lectures (45 %)

- CLT + Random vectors (today)
- Multi-variate Gaussians (next class)
- Markov Chains (2 lectures)
- Statistics

Towards CLT

- Recall $\hat{\mu}_n = \frac{S_n}{n}$ where $S_n = \sum_{i=1}^{n} X_i$
- $\{X_i\}$ is i.i.d. with mean μ amnd variance σ^2.
- $E[\hat{\mu}_n] = \mu$ and $var(\hat{\mu}_n) = \frac{\sigma^2}{n}$
- Now consider $Y_n = \frac{\hat{\mu}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$. (centering and scaling). What is the mean and variance of Y_n?
- $E[Y_n] = 0$ and $Var(Y_n) = 1$. What is $F_{Y_n}(\cdot)$?
- What is $\lim_{n\to\infty} F_{Y_n}(\cdot)$? ANS: $\Phi(\cdot) = F_{N(0,1)}(\cdot)$
- In other words, Y_n converges to $Y = N(0,1)$ in distribution.

CLT

Let $\{X_n, n \geq 0\}$ denote a sequence of i.i.d random variables each with mean μ and variance $0 < \sigma^2 < \infty$.Denote $\hat{\mu}_n = \frac{\sum_{i=1}^{n} X_i}{n}$ and $Y_n = \frac{\hat{\mu}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$. Then Y_n converges to $N(0,1)$ in distribution.

- X_i could be ANY discrete or continuous r.v. with finite mean and variance.
- What is the consequence when $E[X_i] = 0$ and $Var(X_i) = 1$.
- In this case, $Y_n = \frac{S_n}{\sqrt{n}}$ and it converges in distribution to $N(0,1)$.
- $\frac{S_n}{n}$ converges almost surely to 0 but $\frac{S_n}{\sqrt{n}}$ converges to a random variable $\mathcal{N}(0,1)$.

CLT

Let $\{X_n, n \geq 0\}$ denote a sequence of i.i.d random variables each with mean μ and variance $0 < \sigma^2 < \infty$.Denote $\hat{\mu}_n = \frac{\sum_{i=1}^{n} X_i}{n}$ and $Y_n = \frac{\hat{\mu}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$. Then Y_n converges to $N(0,1)$ in distribution.

- CLT given a way to find approximate disribution of $\hat{\mu}_n$.
- Note that for large enough n, we can use the approximation that $Y_n \sim \mathcal{N}(0,1)$.
- Since Gaussianity is preserved under affine transformation, $\hat{\mu}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

Example from probabilitycourse.com

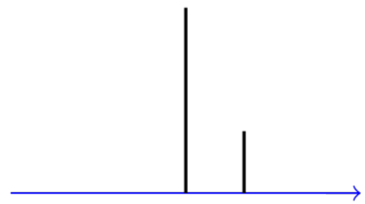
Assumptions:

- $X_1, X_2 \ldots$ are iid Bernoulli(p).
- $Z_n = \dfrac{X_1 + X_2 + \ldots + X_n - np}{\sqrt{np(1-p)}}$.

We choose $p = \frac{1}{3}$.

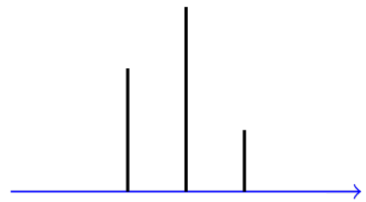
$$Z_1 = \frac{X_1 - p}{\sqrt{p(1-p)}}$$

PMF of Z_1



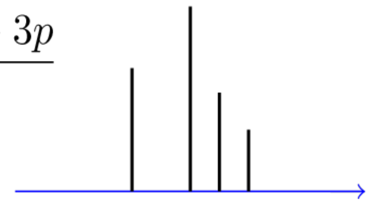
$$Z_2 = \frac{X_1 + X_2 - 2p}{\sqrt{2p(1-p)}}$$

PMF of Z_2



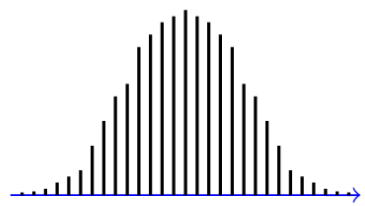
$$Z_3 = \frac{X_1 + X_2 + X_3 - 3p}{\sqrt{3p(1-p)}}$$

PMF of Z_3



$$Z_{30} = \frac{\sum_{i=1}^{30} X_i - 30p}{\sqrt{30p(1-p)}}$$

PMF of Z_{30}



Normal Approximation based on CLT

- Let $S_n = X_1 + \ldots X_n$ where X_i are i.i.d. with mean μ and variance σ^2. If n is large, CDF of S_n can be approximated as follows.

$$P(S_n < c) \approx \Phi(z) \text{ where } z = \frac{c - n\mu}{\sigma\sqrt{n}}$$

https://www.youtube.com/watch?v=zeJD6dqJ5lo&t=111s

Random Vectors

Random Vectors

- We are now moving from a univariate random variable to multivariate random variables, also called as random vectors.
- An n-dimensional random vector is a column vector $\mathbf{X} = (X_1, \ldots X_n)^T$ whose components X_i are scalar valued random variables defined on the same space $(\Omega, \mathcal{F}, P)$.
- Since the components are on the same space, they may be correlated with each other.
- Example:$\mathbf{X} = (X_1, X_2)^T$ where $X_1 = Z_1$ ans $X_2 = Z_1 + Z_2$ where Z_1 and Z_2 are independent standard normal.
- What is the pdf, cdf, marginals, mean, variance/covariance of X?

Random Vectors - Notation

- The CDF and pdf of the random vector $\mathbf{X}$ is denoted as follows :

$$F_{\mathbf{X}}(\mathbf{x}) = F_{X_1,\ldots X_n}(x_1, \ldots x_n)$$

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1,\ldots X_n}(x_1, \ldots x_n)$$

- The joint CDF/pdf captures the correlation between components.

- The expected value vector $E[\mathbf{X}] = (E[X_1], \ldots, E[X_n])^T$

- Linearity of expectation hold here and so for any deterministic matrix $\mathbf{A}$ and vector $\mathbf{b}$ and $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ we have

$$E[\mathbf{Y}] = \mathbf{A}E[\mathbf{X}] + \mathbf{b}.$$

Covariance matrix

- The covariance matrix $C_{\mathbf{X}}$ captures the covariance between components and is defined by

$$
\begin{aligned}
C_{\mathbf{X}} &= E[(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])^T] \\
&= \begin{bmatrix}
Var(X_1) & Cov(X_1, X_2) & \dots & Cov(X_1, X_n) \\
Cov(X_2, X_1) & Var(X_2) & \dots & Cov(X_2, X_n) \\
\vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots \\
Cov(X_n, X_1) & Cov(X_n, X_2) & \dots & Var(X_n)
\end{bmatrix}
\end{aligned}
$$

Covariance matrix: Properties

- The covariance matrix $C_{\mathbf{X}}$ is always positive semi-definite, i.e., for any vector $a \neq 0$ we have $a^T C_{\mathbf{X}} a \geq 0$. Why ?

  Let $u = a^T(\mathbf{X} - E[\mathbf{X}])$, then $a^T C_{\mathbf{X}} a = E[uu^T] = E[u^2] \geq 0$
- If $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$, show that $C_{\mathbf{Y}} = AC_{\mathbf{X}}A^T$. (HW)
- Now recall how we obtained the pdf of Y from pdf of X when $Y = g(X)$

  > Consider $Y = g(X)$ where g is monotone, continuous, differentiable. Then $f_Y(y) = f_X(h(y))|\frac{dh}{dy}(y)|$ where h is the inverse function of g.

- How does this generalize to $\mathbf{Y} = G(\mathbf{X})$? How do we get $f_{\mathbf{Y}}$ from $f_{\mathbf{X}}$?

Functions of random vectors

- Let $\mathbf{Y} = G(\mathbf{X})$ where $G : \mathbb{R}^n \to \mathbb{R}^n$, continuous invertible with continuous partial derivatives.
- Then one can write $\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} G_1(X_1, \dots X_n) \\ G_2(X_1, \dots X_n) \\ \vdots \\ G_n(X_1, \dots, X_n) \end{bmatrix}$
- For example if $\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} 2X_1 \\ X_1 + X_2 \end{bmatrix}$ then $G_1(X_1, X_2) = 2X_1$ and $G_2(X_1, X_2) = X_1 + X_2$.
- What does continuity of G mean? Continuity of components?

Functions of random vectors

- Let H denote inverse of G. We similarly have
$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_n \end{bmatrix} = \begin{bmatrix} H_1(Y_1, \dots Y_n) \\ H_2(Y_1, \dots Y_n) \\ \vdots \\ H_n(Y_1, \dots, Y_n) \end{bmatrix}$$
- For the example we have $X_1 = H_1(Y_1, Y_2) =$ and $X_2 = H_2(Y_1, Y_2) = Y_2 - \frac{Y_1}{2}$.

Functions of random vectors

Let $\mathbf{Y} = G(\mathbf{X})$ where $G : \mathbb{R}^n \to \mathbb{R}^n$, continuous invertible with continuous partial derivatives. Let H denote its inverse. Then

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(H(\mathbf{y}))|J|$$

where J is the determinant of the Jacobian matrix given by

$$\begin{bmatrix} \frac{\partial H_1}{\partial y_1} & \frac{\partial H_1}{\partial y_2} & \cdots & \frac{\partial H_1}{\partial y_n} \\ \frac{\partial H_2}{\partial y_1} & \frac{\partial H_2}{\partial y_2} & \cdots & \frac{\partial H_2}{\partial y_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial H_n}{\partial y_1} & \frac{\partial H_n}{\partial y_2} & \cdots & \frac{\partial H_n}{\partial y_n} \end{bmatrix}$$

Jacobian determinant

- From Vector Calculus: The Jacobian gives the ratio of the incremental areas $dx_1 dx_2..dx_n$ and $dy_1, \ldots dy_n$.
- https://en.wikipedia.org/wiki/Jacobian_matrix_and_determinant
- https://www.khanacademy.org/math/multivariable-calculus/multivariable-derivatives/jacobian/v/jacobian-prerequisite-knowledge
- HW1: For the running example, find $f_{\mathbf{Y}}(y)$.
- HW2: When $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$, how that

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|det(A)|} f_{\mathbf{X}}(A^{-1}(\mathbf{y} - \mathbf{b}))$$