Section 1
EE278: Introduction to Statistical Signal Processing (Fall 2020)

Monday, Sep 21, 2020 - 3 to 4 pm

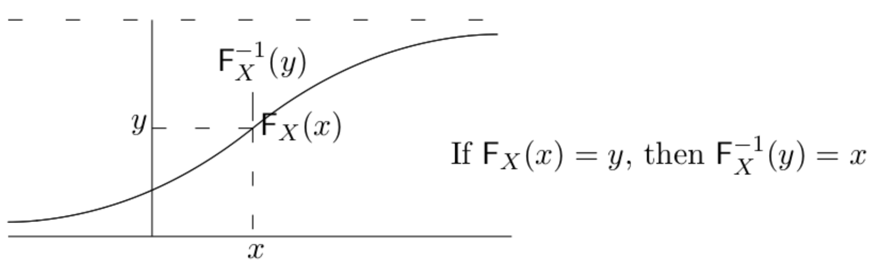
1. Distribution tilting

(a) If X is a continuous random variable having cumulative distribution function (CDF) F, show that the random variable Y defined to be $F(X)$ is uniformly distributed in $(0, 1)$.

(b) Use this to come up with a method to sample a random variable X with CDF F if you can only sample a random variable Y which is uniformly distributed in $(0, 1)$.

Solution:

(a) For a continuous random variable, the CDF is a continuous function. Assume also that $F(x)$ is strictly increasing in x. Then $F(x)$ has an inverse function $F^{-1}(y)$ for $y \in (0, 1)$.



As seen from the plot above, the event $\{F(X) \leq y\}$ is the same as the event $\{X \leq F^{-1}(y)\}$. One can compute the CDF of the random variable $F(X)$ as follows

$$\forall y \in (0,1)\,, \quad \mathbb{P}\left(F(X) \leq y\right) = \mathbb{P}\left(X \leq F^{-1}\left(y\right)\right) = F\left(F^{-1}\left(y\right)\right) = y,$$

where the first equality is due to the monotonicity of F and the second equality follows from the definition of CDF. This coincides with the CDF of $\mathsf{Unif}\,(0, 1)$, completing the proof.

(b) Conversely, suppose Y has a uniform distribution on $(0, 1)$. Then the random variable $X = F^{-1}(Y)$ has the CDF F. This is because

$$\mathbb{P}\left(F^{-1}(Y) \leq x\right) = \mathbb{P}\left(Y \leq F(x)\right) = F(x)$$

Note: The function F^{-1} is also called the quantile function or quantile transformation. It is very useful in sampling random variables with any given CDF using samples from a uniform random variable. In case $F(x)$ is not strictly increasing, we can define $F^{-1}(y)$ to be the generalized inverse distribution function, i.e.

$$F^{-1}\left(y\right) = \inf_x \left\{x : F(x) \geq y\right\}.$$

Then the above properties can still be satisfied.

2. Law of Large Numbers for the Median

In class, we have seen that the LLN shows convergence of the sample average to the mean. Here we will see that the LLN allows us to estimate the entire CDF $F_X(x)$ of a random variable X, and the median of X using several independent samples.

(a) Let $X_1, \ldots, X_n$ be i.i.d. random variables with the CDF $F_X(x)$. For any given y, let $\mathbb{I}_j(y)$ be the indicator function for the event $\{X_j \le y\}$, i.e.

$$\mathbb{I}_j(y) = \begin{cases} 1 & X_j \le y \\ 0 & \text{otherwise} \end{cases}$$

Note that $\mathbb{I}_1(y), \ldots, \mathbb{I}_n(y)$ are also i.i.d. random variables. State the law of large numbers for the random variables $\mathbb{I}_j(y)$.

(b) Suggest a method to estimate the median of X using the samples $X_1, \ldots, X_n$. Assume that X is a continuous random variable and that its PDF is positive in an open interval around the median (Define the median as any value α such that $F_X(\alpha) = 0.5$).

Solution:

(a) The random variable $\mathbb{I}_j(y)$ has the following mean and variance:

$$\mathbb{E}\left[\mathbb{I}_j(y)\right] = \mathbb{P}\left(X_j \le y\right) = F_X(y)$$
$$\text{Var}\left(\mathbb{I}_j(y)\right) = \mathbb{E}\left[\mathbb{I}_j(y)^2\right] - \mathbb{E}\left[\mathbb{I}_j(y)\right]^2 = F_X(y) - F_X(y)^2$$

Applying Chebyshev's inequality to the random variable $\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_j(y)$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_j(y) - F_X(y)\right| > \epsilon\right) \le \frac{\text{Var}\left(\mathbb{I}_j(y)\right)}{n\epsilon^2}$$

Thus the law of large numbers holds, and is as follows:

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_j(y) - F_X(y)\right| > \epsilon\right) = 0 \quad \text{for all } y \text{ and } \epsilon > 0$$

Thus if n is large enough, $\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_j(y)$, which is the fraction of indices j with $X_j \le y$, becomes close to the CDF $F_X(y)$. Further, this LLN holds irrespective of whether X has a finite mean and variance since $\mathbb{I}_j(y)$ always has a finite mean and variance for all y.

(b) The median of a continuous random variable X is any value α such that $F_X(\alpha) = 0.5$. Since $\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_j(x)$ is a good estimator for $F_X(x)$, we can estimate the median as any value $\widehat{\alpha}_n$ for which $\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_j(\widehat{\alpha}_n)$ is the closest to 0.5. Due to the law of large numbers shown above, $\widehat{\alpha}_n$ will be close to α with high probability. This shows that the sample median of n i.i.d. samples converges to the true median if n is large.

A formal proof of the convergence of the sample median to the true median (optional) :

Let $\delta > 0$ be arbitrary. Note that if $\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_j(\alpha - \delta) < 0.5$, then $\widehat{\alpha}_n \ge \alpha - \delta$. Similarly, if $\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_j(\alpha + \delta) > 0.5$, then $\widehat{\alpha}_n \le \alpha + \delta$. Then,

$$\begin{aligned}\mathbb{P}\left(|\widehat{\alpha}_n - \alpha| > \delta\right) &= \mathbb{P}\left(\widehat{\alpha}_n < \alpha - \delta\right) + \mathbb{P}\left(\widehat{\alpha}_n > \alpha + \delta\right) \\ &\le \mathbb{P}\left(\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_j(\alpha - \delta) \ge 0.5\right) + \mathbb{P}\left(\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_j(\alpha + \delta) \le 0.5\right)\end{aligned}$$

Since we assumed that the PDF of X is positive in an interval around α, there exist ϵ_1 such that $F_X(\alpha - \delta) < 0.5 - \epsilon_1$, and ϵ_2 such that $F_X(\alpha + \delta) > 0.5 + \epsilon_2$. Thus,

$$\mathbb{P}\left(|\widehat{\alpha}_n - \alpha| > \delta\right) \le \mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_j(\alpha - \delta) - F_X(\alpha - \delta)\right| > \epsilon_1\right) + \mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_j(\alpha + \delta) - F_X(\alpha + \delta)\right| > \epsilon_2\right)$$

Due to the law of large numbers shown in (a), both the probabilities on the right go to 0 as $n \to \infty$. The assumption that the PDF is positive around α is so that the median is unique. Such a convergence can also be shown without this assumption, by allowing α to be any median.

3. Fun with Gaussian random variables.

Let $X_1, X_2, X_3 \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$. Compute the following quantities:

(a) The distribution of $Y_1 = 2X_1 + 3$.

(b) The distribution of $Y_2 = X_1 + 2X_2$ and the distribution of $Y_3 = X_1 - X_3$.

(c) The correlation between Y_2 and Y_3:

$$\mathrm{corr}(Y_2, Y_3) = \frac{\mathrm{Cov}(Y_2, Y_3)}{\sqrt{\mathrm{Var}(Y_2)\mathrm{Var}(Y_3)}}.$$

Solution.

(a) $Y_1 \sim \mathcal{N}(3, 4)$.

(b) $Y_2 \sim \mathcal{N}(0, 5)$, $Y_3 \sim \mathcal{N}(0, 2)$.

(c) $\mathrm{Cov}(Y_2, Y_3) = \mathbb{E}[Y_2 Y_3] = 1$, $\mathrm{Var}(Y_2) = 5$ and $\mathrm{Var}(Y_3) = 2$. Hence $\mathrm{corr}(Y_2, Y_3) = \frac{1}{\sqrt{10}}$.

4. Eigenvalues and eigenvectors of a symmetric matrix

$$A = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{bmatrix}.$$

(a) Without computing the eigenvalues, what is the sum and product of the eigenvalues of A?

(b) Now compute the eigenvalues of A.

Solution.

(a) Sum of the eigenvalues = $\mathrm{tr}(A) = 6$. Product of the eigenvalues = $\det(A) = 8$.

(b) To calculate the eigenvalues we need to set $\det(A - \lambda I) = 0$:

$$|A - \lambda I| = \begin{vmatrix} 3-\lambda & 2 & 4 \\ 2 & 0-\lambda & 2 \\ 4 & 2 & 3-\lambda \end{vmatrix} = -\lambda^3 + 6\lambda^2 + 15\lambda + 8 = -(\lambda - 8)(\lambda + 1)^2,$$

with roots (eigenvalues) $\lambda_1 = -1$, $\lambda_2 = -1$, $\lambda_3 = 8$.

References

[1] R. G. Gallager. *Stochastic processes: theory for applications.* Cambridge University Press, 2013.