Lecture 28

niceguy

March 22, 2023

1 Review Lecture

1.1 Random Variables

X is a random variable with a set of possible values. It has a density function $f(x)$. In the discrete case, $f(x) = P(X = x)$. In the continuous case, $P(X = x) = 0 \forall x$. However, we have that

$$P(a \leq X \leq b) = \int_a^b f(x) dx \forall a, b \in \mathbb{R}$$

If we have X, Y with distributions $f(x), g(y)$, and $Z = X + Y$. Then

$$h(z) = \sum_{k=-\infty}^{\infty} f(k) g(z-k)$$

In the continuous case,

$$h(z) = \int_{-\infty}^{\infty} f(t) g(z-t) dz$$

Now expectation is linear, i.e.

$$E[aX + bY] = aE[X] + bE[Y]$$

For a change of basis, where X has distribution $f(x)$ and $W = u(X)$.

Then

$$
\begin{aligned}
G(w) &= P(W < w) \\
&= P(X < u^{-1}(w)) \\
&= \int_{-\infty}^{u^{-1}(w)} f(x)dx \\
g(w) &= \frac{dG(w)}{dw} \\
&= f(u^{-1}(w)) \left| \frac{du^{-1}(w)}{dw} \right|
\end{aligned}
$$

Variance is defined as

$$\mathrm{var}(X) = E[(X-\mu)^2]$$

where constants can be taken out and squared, i.e.

$$\mathrm{var}(aX) = a^2\mathrm{var}(X)$$

Now, the variance of a sum is

$$
\begin{aligned}
\mathrm{var}(X+Y) &= E[(X+Y-\mu_x-\mu_y)^2] \\
&= \mathrm{var}(X) + \mathrm{var}(Y) + 2E[(X-\mu_X)(Y-\mu_Y)] \\
&= \sigma_X^2 + \sigma_Y^2 + 2\mathrm{cov}(X,Y)
\end{aligned}
$$

X and Y are uncorrelated if their covariance is 0. If two variables are independent, they are uncorrelated, but the opposite doesn't always hold.

1.2 Sampling

We have random variables $X_1, \ldots, X_n$, which are IID. We also have observations $x_1, \ldots, x_n$. There is a distribution $f(x)$, but we do not know that in general.

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

An unbiased estimator, for example $\overline{X}$, satisfies

$$E[\overline{X}] = \mu$$

The variance is then

$$\begin{aligned}
\mathrm{var}(\overline{X}) &= \mathrm{var}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) \\
&= \frac{1}{n^2}\mathrm{var}\left(\sum_{k=1}^{n} X_k\right) \\
&= \frac{1}{n^2}\sum_{k=1}^{n} \mathrm{var}(X_k) \\
&= \frac{1}{n^2}\sum_{k=1}^{n} \sigma^2 \\
&= \frac{\sigma^2}{n}
\end{aligned}$$

where we use the fact that $\mathrm{var}(X_i + X_j) = \mathrm{var}(X_i) + \mathrm{var}(X_j)$ as X_i, X_j are independent.

1.3 Centre Limit Theorem

Define the statistic

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

The theorem states that despite the distribution, Z tends to a standard normal distribution as $n \to \infty$. We usually take $n \geq 30$ as a threshold. If we define

$$z_{\frac{\alpha}{2}} = -\Phi^{-1}\left(\frac{\alpha}{2}\right)$$

then we get a $1 - \alpha$ confidence interval

$$\left[\overline{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$$