Probability and Statistics Notes

January 15, 2024

1 Sum of random variables

Often we want to find the distribution of sums of two independent r.v $U = X + Y$. There are a few approaches

1.1 PDF change of variables

This approach is based on the probability "volume" being equal after variable change (e.g. $f(x)dx = g(u)du \rightarrow g(u) = f(x(u)) \left|\frac{dx}{du}\right|$)We can choose another $V = X$, then

$$g(u, v) = f(x(u, v), y(u, v)) \left|\frac{\partial(x, y)}{\partial(u, v)}\right|$$

. So the pdf of U is $g(u) = \int g(u, v)dv$. Note that the choice of V is important as sometimes it will make $\int g(u, v)dv$ go unbounded (e.g. when X and Y are i.i.d exponential r.v, the choice of $V = X$ will make the integral infinite).

1.2 Find CDF

When it is hard to choose V, we can try finding the CDF of U directly.

$$P(U = X + Y < u) = \int \int_{(x,y)\in Q} f(x, y)dxdy$$

where $Q = \{(x, y)|x + y < u\}$. Note that $x + y = u$ is a line with negative slope that intercept x and y axis at u. Then we can describe the region of Q and translate it into the integration bound.

1.3 Compute MGF

If X and Y are two r.v and for all t, $M_X(t) = M_Y(t)$ then $F_X(x) = F_Y(x)$ for all values of x (X, Y have the same distribution). We can use this property to calculate the MGF of $X + Y$ and compared it with known MGF of other distributions.

2 Hypothesis testing

2.1 Concepts

1. To study a population, we formulate hypothesis about the parameters H_0 and H_1.
2. We create test (in this case $LRT(x) = \lambda(x)$) to test the hypothesis by inputting the data into the test.
3. We can compute power function of the test $P_\theta(X \in R) = P_\theta(\lambda(x) < c)$ for some c (measures type I error). Note that the c controls the rejection region R here.

4. We define the level/size of the test α to be

$$\sup_{\theta \in \Theta_0} P_\theta(X \in R) = \sup_{\theta \in \Theta_0} P_\theta(\lambda(x) < c) = \alpha \quad (\text{ or } \leq \alpha)$$

   we usually choose $\alpha = 0.05$ to suppress type I error to 5%

5. As we can see, c is related to α, we can find c by solving the above inequality.

6. Once we have c, the rejection region is determined, we can input data into $\lambda(x)$ and test it against c. If $\lambda(x) < c$, we reject H_0.

7. Test is chosen to minimize both type I and type II errors ($P_\theta(X \in R)$ and $P_\theta(X \in R^c)$)