Problem set 5 - Important Distributions in Statistics

Poisson distribution: $X \sim \text{Poisson}(\lambda)$, where X shows the number of occurrences in Poisson process with an average number of events λ.

$$P(X = k) = \frac{e^{-k}\lambda^k}{k!}, \quad k \in \mathbb{N}_0, \quad E(X) = \lambda, \quad Var(X) = \lambda$$

Or if r is an average rate of occurrences, and t is a time period, then:

$$P(X = k) = \frac{e^{-rt}(rt)^k}{k!}, \quad k \in \mathbb{N}_0, \quad E(X) = rt, \quad Var(X) = rt$$

1. Cars independently pass a point on a busy road at an average rate of 150 per hour.
   (a) Assuming a Poisson distribution, find the probability that none passes in a given minute.
   (b) What is the expected number passing in two minutes?
   (c) Find the probability that the expected number actually passes in a given two-minute period.

   Other motor vehicles (vans, motorcycles etc.) pass the same point independently at the rate of 75 per hour. Assume a Poisson distribution for these vehicles too.

   (d) What is the probability of one car and one other motor vehicle in a two-minute period?

$X \sim \text{Exp}(\lambda)$, where X shows the time between events in a Poisson process with average number of events λ.

$$f_X(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & x \geq 0 \end{cases}, F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - \lambda e^{-\lambda x}, & x \geq 0 \end{cases}, \quad E(X) = \frac{1}{\lambda}, \mathsf{Var}(X) = \frac{1}{\lambda^2}$$

2. Telephone calls enter the hotel switchboard according to a Poisson process on the average of two every 3 minutes. Let Y be the waiting time for the second call.
   (a) Calculate the probability, that there will be exactly one call in 3 minutes.
   (b) Calculate the probability, that there will be no more than one call in 3 minutes.
   (c) Find the probability $P(Y > 3)$.
3. Prove the memoryless property of an exponential random variable $X \sim \text{Exp}(\lambda)$:

$$P(X > s + t | X > s) = P(X > t) \quad \forall t, s > 0.$$

Sum of normal distributions

All sums (linear combinations) of independent normally distributed random variables are also normally distributed.

Suppose $X_1, \ldots, X_n$ are independent normally distributed random variables, with $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, \ldots, n$, and $a_1, \ldots, a_n$ and b are constants. Then:

$$S = \sum_{i=1}^{n} a_i X_i + b \sim N(\mu, \sigma^2), \quad \text{where}$$

$$\mu = \sum_{i=1}^{n} a_i \mu_i + b \text{ and } \sigma^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2 + 2 \sum \sum_{i<j} a_i a_j \mathsf{cov}(X_i, X_j)$$

4. Suppose that in the population of Russian people aged 16 or over:

   The heights of men (in cm) follow a normal distribution with mean 174.9 and standard deviation 7.39.

   The heights of women (in cm) follow a normal distribution with mean 161.3 and standard deviation 6.85.

   We select one man and one woman at random and independently of each other. What is the probability that the man is at most 10 cm taller than the woman?

χ^2 distribution

Let $Z_1, Z_2, \ldots, Z_k$ be independent $N(0,1)$ random variables. If:

$$Q = Z_1^2 + Z_2^2 + \ldots + Z_k^2 = \sum_{i=1}^{k} Z_i^2$$

the distribution of Q is the χ^2 distribution with k degrees of freedom. This is denoted by $Q \sim \chi^2(k)$ or $Q \sim \chi^2_k$. The χ^2_k distribution is a continuous distribution, which can take values of $x \geq 0$. Its mean and variance are:

$$E(Q) = k$$

$$Var(Q) = 2k$$

If $Q_1, \ldots, Q_m$ are independent and distributed as $\chi^2_{k_1}, \ldots \chi^2_{k_m}$ respectively, then:

$$Q_1 + \ldots + Q_m \sim \chi^2_{k_1 + \ldots + k_m}$$

5. Let $X_1, \ldots, X_8$ be i.i.d. from $N(0,16)$, $Y_1, \ldots, Y_5$ be i.i.d. from $N(1,9)$; $Z_1, \ldots, Z_4$ be i.i.d. from $N(0,1)$, and all X_i, Y_j, Z_k are independent.

   (a) Find $P\left(\sum_{i=1}^{4} Z_i^2 < 5\right)$

   (b) Find k such that $P\left(\sum_{i=1}^{4} Z_i^2 < k\right) = 0.4$

   (c) Find $P\left(\sum_{i=1}^{8} X_i^2 < 20\right)$

   (d) Find $P\left(\sum_{i=1}^{5} (Y_i - 1)^2 > 32\right)$

t-distribution

Let $Z \sim N(0,1)$ and $Q \sim \chi^2_k$ be independent random variables. Then the the distribution of random variable

$$T = \frac{Z}{\sqrt{Q/k}}$$

is t-distribution with k degrees of freedom (Student's t distribution)

$$E(T) = 0, k > 1$$

$$Var(T) = \frac{k}{k-2}, k > 2$$

When k increases t-distribution goes to normal distribution.

6. Let $X_1, \ldots, X_8$ be i.i.d. from $N(0,16)$, $Y_1, \ldots, Y_5$ be i.i.d. from $N(1,9)$, and all X_i, Y_j are independent.

   (a) Find $P\left(X_1 < \sqrt{\sum_{i=1}^{5} (Y_i - 1)^2}\right)$

   (b) Find $P\left(X_1 + 2X_2 < \sqrt{\sum_{i=1}^{4} (Y_i - 1)^2}\right)$

   (c) Find $P\left(Y_1 - 1 < \sqrt{\sum_{i=1}^{8} X_i^2}\right)$

F-distribution

Let $Q \sim \chi^2_k$ and $R \sim \chi^2_p$ be independent random variables. Then

$$F = \frac{Q/k}{R/p} \sim F_{k,p}$$

is F-distributed with degrees of freedom k and p.
By composition, inverse of F:

$$\frac{1}{F} \sim F_{p,k}$$

If $T \sim t_k$, then $T^2 \sim F_{1,k}$.

$$E(F) = \frac{k}{k-2}, k > 2$$

$$Var(F) = \frac{2k^2(p+k-2)}{p(k-2)^2(k-4)}, k > 4$$

7. Let $X_1, \ldots, X_8$ be i.i.d. from $N(0, 16)$, $Y_1, \ldots, Y_5$ be i.i.d. from $N(1, 9)$, and all X_i, Y_j are independent.

   (a) Find $P\left(\sum_{i=1}^{8} X_i^2 < \sum_{i=1}^{5} (Y_i - 1)^2\right)$

   (b) Find $P\left(\sum_{i=1}^{4} X_i^2 < 7\sum_{i=1}^{3} (Y_i - 1)^2\right)$

   (c) Find $P\left(\sum_{i=1}^{3} X_i^2 < \sum_{i=4}^{8} X_i^2\right)$

   (d) Find $P\left(\sum_{i=1}^{7} X_i^2 < \sum_{i=3}^{8} X_i^2\right)$

8. Suppose that X_1 and X_2 are independent $N(0, 4)$ random variables. Compute $P(X_1^2 < 36.84 - X_2^2)$.

9. Suppose that X_1, X_2 and X_3 are independent $N(0, 1)$ random variables, while Y (independently) follows a χ^2_5 distribution. Compute $P(X_1^2 + X_2^2 < 7.236Y - X_3^2)$.

10. Suppose $X_i \sim N(0, 3)$, for $i = 1, 2, 3, 4$. Assume all these random variables are independent. Derive the value of k in each of the following:

    a) $P(X_1 + 3X_2 > 4) = k$

    b) $P(X_1^2 + X_2^2 + X_3^2 + X_4^2 < k) = 0.99$

    c) $P(X_1 < k\sqrt{X_2^2 + X_3^2}) = 0.05$

    d) $P(X_1^2 + X_2^2 > 39(X_3^2 + X_4^2)) = k$