Actuary Probability I

Expected Values and Moments

Irving Gómez Méndez

August-December, 2021

UNIVERSIDAD IBEROAMERICANA
CIUDAD DE MÉXICO ®

Statistics of a Distribution

Moments, Expected Value

If X is a discrete random variable, the moment of order n of X is given by

$$\mathbb{E}[X^n] = \sum_i x_i^n \mathbb{P}(X = x_i),$$

always that the series converges absolutely. If the series diverges, we say that the moment does not exists.

If X is a continuous random variable with density $f(x)$, the moment of order n of X is given by

$$\mathbb{E}[X^n] = \int_{-\infty}^{\infty} x^n f(x) dx,$$

always that the integral converges absolutely.

The first moment, corresponding to $n = 1$, is known as the mean or expected value of X, denoted by μ.

Central Moments, Variance, Standard Deviation

The central moment of order n is the moment of order n of the variable $X - \mu$, always that μ exists. The nth central moment is denoted as μ_n. The first central moment is zero. The second central moment is called the variance of X, denoted as $\mathbb{V}[X]$. We can show that

$$\mathbb{V}[X] = \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X] - \mu.$$

The squared root of the variance is called the standard deviation of X, denoted by $sd(X)$. It is common to denote the variance of X by σ^2 and the standard deviation by σ.

Coefficient of Variation

The coefficient of variation c_v of a random variable X with mean $\mu \neq 0$ and standard deviation σ is defined as the ratio

$$c_v = \frac{\sigma}{|\mu|}.$$

Empirical results in agriculture studies has shown that when $c_v \leq 0.3$, the mean is “representative” of the random variable, while larger values than 0.3 are translated as an “unrepresentative” mean.

The coefficient of variation is common in applied probability fields such as renewal theory, queueing theory, and reliability theory. These fields consider positive random variables, and hence $\mu > 0$. The standard deviation of an exponential distribution is equal to its mean, so its coefficient of variation is equal to 1. Distributions with $c_v < 1$ (such as an Erlang distribution) are considered low-variance, while those with $c_v > 1$ are considered high-variance.

Median and Mode

The median of a random variable X is any value ν such that

$$\mathbb{P}(X \geq \nu) \geq \frac{1}{2}, \quad \mathbb{P}(X \geq \nu) \geq \frac{1}{2}.$$

If X is a discrete random variable, the mode is the value x at which the probability mass function takes its maximum value. The mode is not necessarily unique to a given discrete distribution, since the probability mass function may take the same maximum value at several points $x_1, x_2, \ldots$ The most extreme case occurs in uniform distributions, where all values occur equally frequently.

When the density function of a continuous distribution has multiple local maxima it is common to refer to all of the local maxima as modes of the distribution. Such a continuous distribution is called multimodal (as opposed to unimodal). A mode of a continuous probability distribution is often considered to be any value x at which its density function has a locally maximum value, so any peak is a mode.

Skewness

The skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. The skewness value can be positive, zero, negative, or undefined. The coefficient of skewness of a random variable X is defined by

$$\tilde{\mu}_3 = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3}$$

Kurtosis

The kurtosis is a measure of the "tailedness" of the probability distribution of a real-valued random variable. It is related to the tails of the distribution, higher kurtosis corresponds to greater extremity of outliers. The kurtosis of a random variable X is defined as

$$\kappa(X) = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{\mu_4}{\sigma^4}.$$

A related quantity is the excess of kurtosis, defined as $\kappa(X) - 3$. Some authors use the excess of kurtosis as the definition of kurtosis. Random variables whose excess of kurtosis is negative are referred as sub-gaussian, whereas radom variables whose excess of kurtosis is positive are referred as super-gaussian.

Quantiles and Percentiles

Let $F : \mathbb{R} \to [0, 1]$ be the distribution function of a random variable X, we define the quantile function $Q : (0, 1) \to \mathbb{R}$ as the generalized inverse of F, that is

$$Q(p) = \inf\{x, \text{ s.t. } F(x) \geq p\}.$$

The quantile x of probability p of the distribution is such that

$$Q(p) = x$$

or, equivalently as the value x such that

$$F(x) = p.$$

The quantile is the lowest value for which the random variable has accumulated a probability p.
The quantiles of probability $p = n/100$, $n = 1, 2, \ldots, 99$ are known as the percentiles of of probability $n \times 100\%$.

Summary

- Central tendency statistics:
  - Mean or expected value
  - Median
  - Mode
- Dispersion statistics:
  - Variance
  - Standard deviation
  - Coefficient of variation
- Shape statistics:
  - Skewness coefficient
  - Kurtosis coefficient
- Location statistics:
  - Quantiles (percentiles)

Expected Value of a Function of a R.V.

If X is a random variable and g is a function then $g(X)$ is also a random variable. If X is discrete taking values x_j, $j = 1, 2, \dots$ then the expected value of $g(X)$ is given by

$$\mathbb{E}[g(X)] = \sum_j g(x_j)\mathbb{P}(X = x_j),$$

always that the sum converges absolutely. IfX is continuous with density f, the expected value of $g(X)$ is given by

$$\mathbb{E}[g(X)] = \int g(x)f(x)dx.$$

Markov's, Chebyshev's and Jensen's Inequalities

Markov's Inequality

If X is a non-negative random variable ($X \geq 0$), then for any $t > 0$,

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$

Proof

We are going to present two proofs of Markov's inequality, the first one for continuous random variables and the second the proof for continuous or discrete random variables.

(a.) If X has density f,

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty xf(x)dx = \int_0^t xf(x)dx + \int_t^\infty xf(x)dx \\
&\geq \int_t^\infty xf(x)dx \geq t\int_t^\infty f(x)dx \\
&\geq t\mathbb{P}(X \geq t).
\end{aligned}$$

(b.) Since X is non-negative, then for all $t > 0$
$X\mathbb{1}_{X \geq t} \geq t\mathbb{1}_{X \geq t}$, taking the expected value on both sides

$$\begin{aligned}
&\mathbb{E}[t\mathbb{1}_{X \geq t}] \leq \mathbb{E}[X\mathbb{1}_{X \geq t}] \\
\Leftrightarrow &\mathbb{E}[\mathbb{1}_{X \geq t}] \leq \frac{\mathbb{E}[X\mathbb{1}_{X \geq t}]}{t} \\
\Leftrightarrow &\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X\mathbb{1}_{X \geq t}]}{t} \leq \frac{\mathbb{E}[X]}{t}
\end{aligned}$$

Chebyshev's Inequality

If X is a random variable with mean μ and variance σ^2, for any $t > 0$,

$$\mathbb{P}(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}.$$

Proof

Since $(X - \mu)^2$ is a non-negative random variable, we can apply Markov's inequality

$$\mathbb{P}((X - \mu)^2 \geq t^2) \leq \frac{\mathbb{E}[(X - \mu)^2]}{t^2},$$

but $(X - \mu)^2 \geq t^2$ if and only if $|X - \mu| \geq t$, so

$$\mathbb{P}(|X - \mu| \geq t^2) \leq \frac{\sigma^2}{t^2}.$$

Jensen's Inequality

If X is a random variable and g is a convex function, then

$$g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)],$$

always that both expected values exist and are finite.

Proof

Expanding $g(x)$ in a Taylor's series expansion about $\mu = E[X]$ yields

$$g(x) = g(\mu) + g'(\mu)(x - \mu) + \frac{g''(\xi)(x - \mu)^2}{2}$$

where ξ is some value between x and μ. Since $g''(\xi) \geq 0$, we obtain

$$g(x) \geq g(\mu) + g'(\mu)(x - \mu).$$

Hence,

$$g(X) \geq g(\mu) + g'(\mu)(X - \mu).$$

Taking expectations yields

$$\mathbb{E}[g(X)] \geq g(\mu) + g'(\mu)\mathbb{E}[X - \mu] = g(\mu)$$

and the inequality is established.

Conditional Expectation

Joint Distribution

If we have a pair of random variables (X, Y) defined in a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, their joint distribution function is given by

$$F_{XY}(x, y) = F(x, y)\mathbb{P}(X \leq x, Y \leq y).$$

If both distributions are discrete and take values x_i, $i \geq 1$ and y_j, $j \geq 1$ respectively, their joint probability function is

$$p_{XY}(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j), \quad i \geq 1, j \geq 1.$$

A joint distribution function has (joint) density if exists a function f_{XY} of two variables such that

$$F_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(s, t) dt ds, \quad \text{for all } x, y.$$

Marginal Distributions

If both random variables are discrete, the marginal probability functions are given by

$$p_X(x_i) = \sum_j p_{XY}(x_i, y_j) \quad \text{and} \quad p_Y(y_j) = \sum_i p_{XY}(x_i, y_j).$$

If F has a joint density f, the marginal densities of X and Y are given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)dy \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)dx.$$

If X and Y have joint distribution then

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

when all these moments exist.

Independence

If for any x and y $F(x, y) = F_X(x) \times F_Y(y)$, then X and Y are independent. If the variables are discrete with joint probability p_{XY}, they are independent if and only if

$$p_{XY}(x, y) = p_X(x)p_Y(y).$$

Similarly, if the variables are continuous with joint density $f_{XY}(x, y)$, they are independent if and only if

$$f_{XY}(x, y) = f_X(x)f_Y(y).$$

If X and Y are independent r.v. with first moment finite, then the product XY also has first moment finite and

$$E[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

Covariance

If X and Y are r.v. with joint distribution, means μ_x and μ_Y, and finite variances σ_X^2 and σ_Y^2, the covariance of X and Y, denoted as σ_{XY} or $\mathrm{Cov}(X, Y)$ is defined as

$$\sigma_{XY} = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[XY] - \mu_X \mu_Y.$$

We say that X and Y are uncorrelated if $\sigma_{XY} = 0$.

Independent variables with finite variance are uncorrelated, but there are uncorrelated variables that are not independent.

Example

1. If $X \sim \mathcal{N}(0, 1)$ and $Y = X^2$, then $\text{Cov}(X, Y) = \mathbb{E}[XY] = \mathbb{E}[X^3]$, and it can be shown that third moment of a standard normal distribution is zero. Moreover, all the odd moments of the standard normal distribution are zero.
2. If X and Y are independent with variance σ_X^2 and σ_Y^2 respectively, then the variance of the sum $Z = X + Y$ is the sum of the variances:

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2.$$

Conditional Probability and Conditional Expectation

Let X, Y be discrete random variables. The conditional probability function $p_{X|Y}(x|y)$of X given $Y = y$ is defined by

$$p_{X|Y}(x|y) = \frac{p_{XY}(x,y)}{p_Y(y)} \quad \text{if } p_Y(y) > 0.$$

Let g be a function such that $\mathbb{E}[g(X)] < \infty$. We define the conditional expected value of $g(X)$ given $Y = y$ as

$$\mathbb{E}[g(X)|Y = y] = \sum_x g(x) p_{X|Y}(x|y) \quad \text{if } p_Y(y) > 0.$$

Let X, Y be random variables with joint density $f_{XY}(x, y)$. We define the conditional density

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)} \quad \text{if } f_Y(y) > 0.$$

Let g be a function such that $\mathbb{E}[g(X)] < \infty$. We define the conditional expected value of $g(X)$ given $Y = y$ as

$$\mathbb{E}[g(X)|Y = y] = \int g(x) f_{X|Y}(x|y) dx \quad \text{if } f_Y(y) > 0.$$

Expected Value Conditioned in an Event

Let X, Y be r.v. with joint distribution, and let $I \subset \mathbb{R}$, then

$$\mathbb{E}[X|Y \in I] = \frac{\mathbb{E}[X \mathbb{1}_{Y \in I}]}{\mathbb{P}(Y \in I)}, \quad \text{if } \mathbb{P}(Y \in I) > 0.$$

Proof

(Discrete case.)

$$\begin{aligned}
\mathbb{E}[X|Y \in I] &= \sum_i x_i p_{X|Y}(x_i|Y \in I) \\
&= \sum_i x_i \frac{p_{XY}(x_i, Y \in I)}{p_Y(Y \in I)} \\
&= \frac{\sum_i \sum_{j: y_j \in I} x_i p_{XY}(x_i, y_j)}{\mathbb{P}(Y \in I)} \\
&= \frac{\sum_i \sum_j x_i \mathbb{1}_{y_j \in I} p_{XY}(x_i, y_j)}{\mathbb{P}(Y \in I)} \\
&= \frac{\mathbb{E}[X \mathbb{1}_{Y \in I}]}{\mathbb{P}(Y \in I)}.
\end{aligned}$$

(Continuous case.)

$$
\begin{aligned}
\mathbb{E}[X|Y \in I] &= \int x f_{X|Y}(x|Y \in I)dx \\
&= \int \frac{f_{XY}(x, Y \in I)}{f_Y(Y \in I)}dx \\
&= \frac{\int \int_I x f_{XY}(x,y)dydx}{\mathbb{P}(Y \in I)} \\
&= \frac{\int \int x \mathbb{1}_{y \in I} f_{XY}(x,y)dydx}{\mathbb{P}(Y \in I)} \\
&= \frac{\mathbb{E}[X \mathbb{1}_{Y \in I}]}{\mathbb{P}(Y \in I)}.
\end{aligned}
$$

In particular, if X is discrete with probability function p then

$$\mathbb{E}[X|X \in I] = \frac{\mathbb{E}[X\mathbb{1}_{X\in I}]}{\mathbb{P}(X \in I)} = \frac{\displaystyle\sum_{i:x_i\in I} x_i p(x_i)}{\displaystyle\sum_{i:x_i\in I} p(x_i)},$$

whereas if X is continuous with density function f then

$$\mathbb{E}[X|X \in I] = \frac{\mathbb{E}[X\mathbb{1}_{X\in I}]}{\mathbb{P}(X \in I)} = \frac{\displaystyle\int_I xf(x)dx}{\displaystyle\int_I f(x)dx}.$$

Example

Let $X \sim \text{Exp}(\lambda)$, with density $\lambda e^{-\lambda x} \mathbb{1}_{x \geq 0}$ and let $\tau > 0$, then

$$\mathbb{E}[X|X > \tau] = \frac{\int_\tau^\infty x \lambda e^{-\lambda x} dx}{\mathbb{P}(X > \tau)}$$
$$= e^{\lambda \tau} \int_\tau^\infty \lambda x e^{-\lambda x} dx.$$

Take

$$\begin{array}{ll} u = \lambda x, & du = \lambda dx, \\ dv = e^{-\lambda x} dx, & v = -\frac{1}{\lambda} e^{-\lambda x} \end{array}$$

Doing integration by parts,

$$\mathbb{E}[X|X > \tau] = e^{\lambda \tau} \left[-x e^{-\lambda x} \Big|_\tau^\infty + \frac{1}{\lambda} \int_\tau^\infty \lambda e^{-\lambda x} dx \right]$$
$$= e^{\lambda \tau} \left[\tau e^{-\lambda \tau} + \frac{1}{\lambda} e^{-\lambda \tau} \right]$$
$$= \tau + \frac{1}{\lambda}.$$

Probability Generating Function

Probability Generating Function

Consider a r.v. ξ with non-negative values ad probability distribution

$$\mathbb{P}(\xi = k) = p_k, \quad k = 0, 1, \ldots .$$

The probability generating function (p.g.f.) $\phi(s)$ of the r.v. ξ (or equivalently of its distribution) is defined as

$$\phi(s) = \mathbb{E}\left[s^{\xi}\right] = \sum_{k=0}^{\infty} s^k p_k, \quad 0 \le s \le 1.$$

From the definition is immediate that

$$\phi(1) = \sum_{k=0}^{\infty} p_k = 1.$$

Properties of the p.g.f.

1. It is possible to recover the probabilities p_k from ϕ with the formula
$$p_k = \frac{1}{k!} \left. \frac{d^k \phi(s)}{ds^k} \right|_{s=0}.$$
For example,
$$\phi(s) = p_0 + p_1 s + p_2 s^2 + \cdots \Rightarrow p_o = \phi(s)$$
$$\frac{d\phi(s)}{ds} = p_1 + 2p_2 s + 3p_3 s^2 + \cdots \Rightarrow p_1 = \left. \frac{d\phi(s)}{ds} \right|_{s=0}$$
2. If $\xi_1, \ldots, \xi_n$ are independent r.v. with p.g.f. $\phi_1(s), \phi_2(s), \ldots, \phi_n(s)$, the p.g.f. of the sum $X = \xi_1 + \xi_2 + \cdots + \xi_n$ is the product of the probability generated functions
$$\phi_X(s) = \phi_1(s)\phi_2(s) \cdots \phi_n(s).$$

3. The moments of a random variable that takes values in the non-negative integers can be obtained differentiating the probability generating function:

$$\frac{d\phi(s)}{ds} = p_1 + 2p_2 s + 3p_3 s^2 + \cdots ,$$

hence

$$\left.\frac{d\phi(s)}{ds}\right|_{s=1} = p_1 + 2p_2 + 3p_3 + \cdots = \mathbb{E}[\xi].$$

For the second derivative we have

$$\frac{d^2\phi(s)}{ds^2} = 2p_2 + 3\cdot 2p_3 s + 4\cdot 3p_4 s^2 + \cdots ,$$

Evaluating in $s = 1$,

$$\left.\frac{d^2\phi(s)}{ds^2}\right|_{s=1} = 2p_2 + 3 \cdot 2p_3 + 4 \cdot 3p_4 + \cdots$$
$$= \sum_{k=2}^{\infty} k(k-1)p_k$$
$$= \mathbb{E}[\xi(\xi - 1)] = \mathbb{E}[\xi^2] - \mathbb{E}[\xi]$$

so

$$\mathbb{E}[\xi^2] = \left.\frac{d^2\phi(s)}{ds^2}\right|_{s=1} + \mathbb{E}[\xi] = \left.\frac{d\phi(s)}{ds}\right|_{s=1}.$$

Example

Suppose that $\xi \sim \mathrm{Pois}(\lambda)$:

$$p_k = \mathbb{P}(\xi = k) = \frac{\lambda^k}{k|} e^{-\lambda}, \quad k = 0, 1, \ldots .$$

Its probability generating function is

$$\begin{aligned}
\phi(s) = \mathbb{E}\left[s^{\xi}\right] &= \sum_{k=0}^{\infty} s^k \frac{\lambda^k}{k|} e^{-\lambda} \\
&= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(s\lambda)^k}{k!} = e^{-\lambda} e^{s\lambda} \\
&= e^{-\lambda(1-s)}.
\end{aligned}$$

Hence,

$$\frac{d\phi(s)}{ds} = \lambda e^{-\lambda(1-s)}, \quad \left.\frac{d\phi(s)}{ds}\right|_{s=1} = \lambda$$

$$\frac{d^2\phi(s)}{ds^2} = \lambda^2 e^{-\lambda(1-s)}, \quad \left.\frac{d^2\phi(s)}{ds^2}\right|_{s=1} = \lambda^2$$

so, we have

$$\mathbb{E}[\xi] = \lambda, \quad \mathbb{V}[\xi] = \lambda^2 + \lambda - (\lambda)^2 = \lambda.$$

Moment Generating Function

Moment Generating Function

Given a random variable X, or its distribution function F, we define the moment generating function (m.g.f.) as

$$M(t) = \mathbb{E}[e^{tX}]$$

when this expected value exists.

- If the support of X are the non-negative integers, $M_X(t) = \phi_X(e^t)$.
- If X is bounded, M is defined for all $t \in \mathbb{R}$.
- If X is not bounded, the domain of M might not be $\mathbb{R}$. Int this case, M is always defined in zero and $M(0) = 1$.

If the function M is defined around $t = 0$, then the series

$$M(t) = \mathbb{E}[e^{tX}] = \mathbb{E}\left[1 + \sum_{n=1}^{\infty}\right] \frac{t^n X^n}{n!} = 1 + \sum_{n=1}^{\infty} \frac{t^n}{n!}\mathbb{E}[X^n]$$

converges and it can be differentiated. Hence,

$$M'(0) = \mathbb{E}[X], \quad M''(0) = \mathbb{E}[X^2] \quad \text{in general } M^{(n)}(0) = \mathbb{E}[X^n].$$

Examples

1. If $X \sim \text{Bin}(n, p)$,

$$M(t) = \sum_{j=0}^{\infty} e^{jt} \binom{n}{j} p^j (1-p)^{n-j} = (pe^t + 1 - p)^n.$$

2. If $X \sim \text{Exp}(\lambda)$, that is, $f(x) = \lambda e^{-\lambda x} \mathbb{1}_{x \in [0,\infty)}$ for $x \geq 0$, then $M(t) = \lambda/(\lambda - t)$ for $t < \lambda$.

$$M(t) = \int_0^{\infty} \lambda e^{-\lambda x} e^{tx} dx = \lambda \frac{e^{(t-\lambda)x}}{t-\lambda}\Bigg|_0^{\infty} = \frac{\lambda}{\lambda - t}.$$

Note that $M(t)$ is not defined if $t \geq \lambda$.

3. If $X \sim \mathcal{N}(0,1)$, then

$$\begin{aligned} M(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2 e^{t^2/2}} dx \\ &= e^{t^2/2}. \end{aligned}$$

Since

$$e^t = 1 + t + \frac{t^2}{2} + \frac{t^3}{3!} + \cdots ,$$

$$\begin{aligned} &M(t) = e^{t^2/2} \\ &= 1 + \frac{t^2}{2} + \frac{t^4}{2^2 2!} + \frac{t^6}{2^3 3!} + \cdots \\ &= 1 + 0t + \frac{1}{2}t^2 + 0t^3 + \frac{3}{4!}t^4 + 0t^5 + \cdots \\ &= M(0) + M'(0)t + \frac{M''(0)}{2}t^2 + \frac{M^{(3)}(0)}{3!}t^3 + \frac{M^{(4)}(0)}{4!}t^4 + \cdots . \end{aligned}$$

Thus,

$$\mathbb{E}[X^k] = M^{(k)}(0) = 0, \quad \text{if } k \text{ is odd.}$$

Moreover

$$\begin{aligned}
\mathbb{E}[X] &= M'(0) = 0 && \text{Expected Value,}\\
\mathbb{E}[X^2] &= M''(0) = 1 = \mathbb{V}[X] && \text{Variance,}\\
\mathbb{E}[X^3] &= M^{(3)}(0) = 0 = \tilde{\mu}_3 && \text{Skewness,}\\
\mathbb{E}[X^4] &= M^{(4)}(0) = 3 = \kappa && \text{Kurtosis.}
\end{aligned}$$

De Moivre - Laplace Central Limit

If $S_n \sim \text{Bin}(np)$ for $n \geq 1$. Define $q = 1 - p$ and

$$T_n = \frac{S_n - np}{\sqrt{npq}}.$$

Then for all $x \in \mathbb{R}$,

$$\mathbb{P}(T_n \leq x) \to \Phi(x)$$

Proof
Remember that S_n can be expressed as the sum of n independent and identically distributed (i.i.d.) random variables (r.v.) with Bernoulli distribution of parameter p: $S_n = \sum_{i=1}^{n} X_i$. Then, the moment generating function of T_n is

$$\mathbb{E}\left[e^{tT_n}\right] = \mathbb{E}\left[\exp\left\{\frac{t(S_n - np)}{(npq)^{1/2}}\right\}\right] = \mathbb{E}\left[\exp\left\{\frac{t(\sum_{i=1}^n (X_i - p))}{(npq)^{1/2}}\right\}\right]$$

$$= \mathbb{E}\left[\prod_{i=1}^n \exp\left\{\frac{t(X_i - p)}{(npq)^{1/2}}\right\}\right] \overset{\text{ind.}}{=} \prod_{i=1}^n \mathbb{E}\left[\exp\left\{\frac{t(X_i - p)}{(npq)^{1/2}}\right\}\right]$$

$$= \left(\mathbb{E}\left[\exp\left\{\frac{t(X_1 - p)}{(npq)^{1/2}}\right\}\right]\right)^n \ (X_i \text{ are identically distributed})$$

$$= \left(p\exp\left\{\frac{t(1-p)}{(npq)^{1/2}}\right\} + (1-p)\exp\left\{\frac{-tp}{(npq)^{1/2}}\right\}\right)^n.$$

We write the Taylor's polynomial of the two exponential functions, given by

$$p \exp\left\{\frac{t(1-p)}{(npq)^{1/2}}\right\} = p\left(1 + \frac{qt}{(npq)^{1/2}} + \frac{q^2t^2}{2npq} + \frac{C_1 q^3 t^3}{3!(npq)^{3/2}}\right)$$

$$(1-p) \exp\left\{\frac{-tp}{(npq)^{1/2}}\right\} = q\left(1 - \frac{pt}{(npq)^{1/2}} + \frac{p^2t^2}{2npq} - \frac{C_2 p^3 t^3}{3!(npq)^{3/2}}\right)$$

The sum of the two expressions are

$$1 + \frac{t^2}{2n} + O(n^{-3/2}),$$

hence

$$\mathbb{E}\left[e^{tT_n}\right] = \left(1 + \frac{t^2}{2n} + O(n^{-3/2})\right)^n \to e^{t^2/2}$$

which is the m.g.f. of a standard normal.