Statistics I – Supplements for Chapter 6 More about Continuous Distributions

Ling-Chieh Kung

Department of Information Management
National Taiwan University

October 24, 2012

Introduction

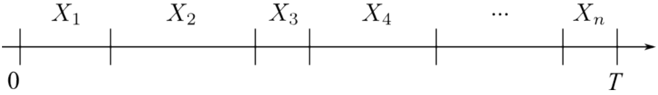
- In this supplements, we will introduce two more continuous probability distributions:
  - The gamma distribution.
  - The chi-square (χ^2) distribution.
- We will also prove Chebyshev's theorem.

Road map

- **Gamma distributions**.
  - Exponential distributions.
  - Chi-square distributions.
- Proof of Chebyshev's theorem.

Sum of interarrival times

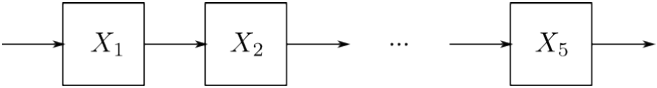
- Consider a consumer arrival process into a store.
  - Let X_0 be the arrival time of the first consumer.
  - Let X_i be the interarrival time between consumers i and $i+1$.
  - It is often assumed that X_is follow an exponential distribution.



- If we only have 10 units to sell, when may we close the store?
  - Let $Y = \sum_{i=1}^{10} X_i$. Then we close at time Y.
  - But what is the distribution of Y?

Sum of service times

- Consider a sequence of services.
  - E.g., a health inspection with 5 steps.
  - Let X_i be the service time at step i, $i = 1, 2, ..., 5$.
  - In some cases, X_is follow an exponential distribution.



- When may we leave the hospital?
  - Let $Y = \sum_{i=1}^{5} X_i$. Then we may leave at time Y.
  - But what is the distribution of Y?

Gamma distributions

- To answer the questions, let's define the **gamma distribution**:

Definition 1 (Gamma distribution)

A random variable X follows the gamma distribution with parameters $\alpha > 0$ and $\beta > 0$, denoted by $X \sim \text{Gamma}(\alpha, \beta)$, if its pdf is

$$f(x|\alpha, \beta) = \frac{x^{\alpha-1} e^{-\frac{y}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)},$$

for all $x \geq 0$, where $\Gamma(\cdot)$ is the gamma function

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx.$$

Exponential distributions redefined

- The gamma distribution itself is not used a lot in modeling practical situations.
- Nevertheless, two of its special cases are used a lot:

Definition 2 (Exponential distribution; alternative)

A random variable X follows the exponential distribution with rate $\lambda > 0$, denoted by $X \sim \text{Exp}(\lambda)$, if it follows the gamma distribution with $\alpha = 1$ and $\lambda = \frac{1}{\beta}$.

Exponential distributions redefined

- So any exponential distribution is a special case of a gamma distribution with $\alpha = 1$.
  - When $\alpha = 1$, $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ becomes $\int_0^\infty e^{-x} dy = 1$.
  - When $\alpha = 1$, $f(x|\alpha, \beta) = \frac{x^{\alpha-1} e^{-\frac{y}{\beta}}}{\beta^\alpha \Gamma(\alpha)}$ becomes $\frac{e^{\frac{-x}{\beta}}}{\beta}$.
  - When $\beta = \frac{1}{\lambda}$, we have $f(x|\lambda) = \lambda e^{-\lambda x}$.
- This is an **alternative** definition of the exponential distribution.

Sum of exponential RVs

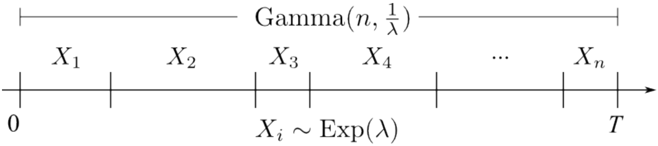
- Now we may answer our original question:

Proposition 1

Let $X_i \sim \text{Exp}(\lambda) \sim \text{Gamma}(1, \frac{1}{\lambda})$. *If* $X_i s$ *are independent, then*

$$\sum_{i=1}^{n} X_i \sim \text{Gamma}\left(n, \frac{1}{\lambda}\right).$$

Proof. Later in this semester. □



Chi-square distributions

- Another special case of the gamma distribution is the **chi-square (χ^2) distribution**.

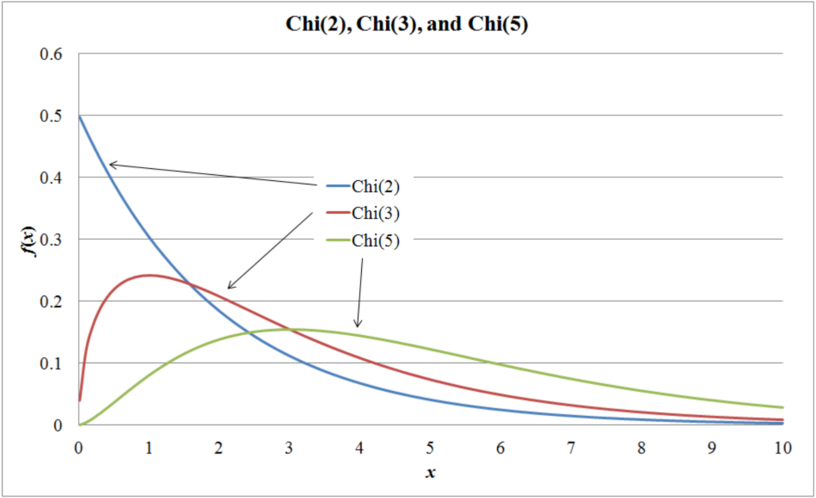
Definition 3 (Chi-square distribution)

A random variable X follows the chi-square distribution with degree of freedom $n \in N$, denoted by $X \sim \chi^2(n)$ or $X \sim \mathrm{Chi}(n)$ if it follows the gamma distribution with $\alpha = \frac{n}{2}$ and $\beta = 2$.

Chi-square distributions

- The chi-square distribution is one of the most important sampling distributions in the world of business Statistics.
- In this semester, it will be used in Chapters 7, 8, and 9.
- Its parameter is called the **degree of freedom**. The reason will be discussed later in this semester.

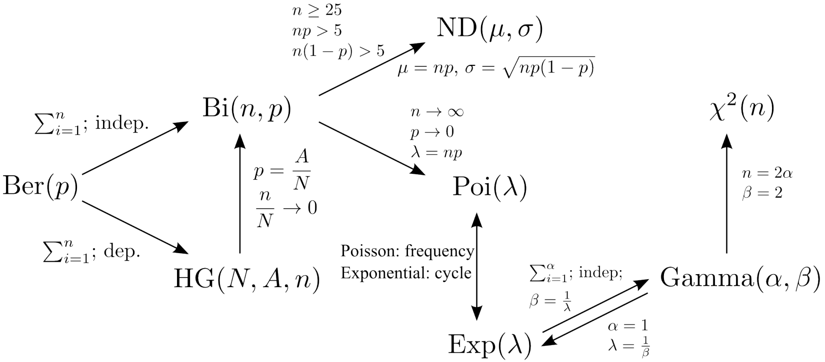
Chi-square distributions



Summary

- The gamma distribution is a continuous distribution.
  - The exponential distribution is a special case of it.
  - The chi-square distribution is a special case of it.
- The sum of independent exponential random variables is a gamma distribution.
- The chi-square distribution will be used extensively in inferential Statistics.

Relationships



Road map

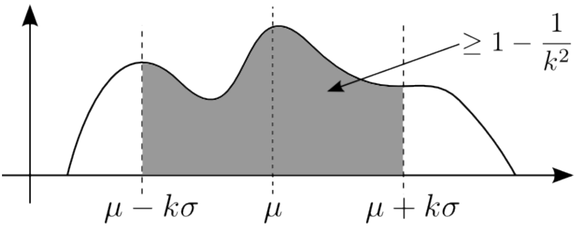
- Gamma distributions.
  - Exponential distributions.
  - Chi-square distributions.
- **Proof of Chebyshev's theorem**.

Chebyshev's theorem for data

- Recall the Chebyshev's theorem for data in Chapter 3:

Proposition 2 (Chebyshev's theorem for data)

For any set of data with mean μ and standard deviation σ, if $k \geq 1$, at least $1 - \frac{1}{k^2}$ proportion of the values are within $[\mu - k\sigma, \mu + k\sigma]$.



Chebyshev's theorem

- In general, the Chebyshev's theorem applies to **any random variable**:

Proposition 3

Let X be a random variable with finite mean μ and variance σ^2. For any $k > 1$,

$$\Pr(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2},$$

or, equivalently,

$$\Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Proof of Chebyshev's theorem

Proof. Here we prove the theorem for a continuous RV. The proof for a discrete RV is similar. Let $f(x)$ be the pdf of X, then

$$\begin{aligned}\mathrm{Var}(X) = \sigma^2 &\equiv \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \\ &= \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)dx + \int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 f(x)dx \\ &\quad + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)dx.\end{aligned}$$

Note that we do this split because we want to study what is happening in the interval $[\mu - k\sigma, \mu + k\sigma]$.

Proof of Chebyshev's theorem

Proof (cont'd). Now, in the first and third integral $(x-\mu)^2 \geq k^2\sigma^2$. Moreover, in the second integral $(x-\mu)^2 \geq 0$. So we may make a substitution and obtain

$$\begin{aligned}
\mathrm{Var}(X) = \sigma^2 &\geq \int_{-\infty}^{\mu-k\sigma} k^2\sigma^2 f(x)dx + \int_{\mu+k\sigma}^{\infty} k^2\sigma^2 f(x)dx \\
&= k^2\sigma^2\left(\int_{-\infty}^{\mu-k\sigma} f(x)dx + \int_{\mu+k\sigma}^{\infty} f(x)dx\right) \\
&= k^2\sigma^2\big[\Pr(X \leq \mu - k\sigma) + \Pr(X \geq \mu + k\sigma)\big] \\
&= k^2\sigma^2\Pr\big(|X-\mu| \geq k\sigma\big).
\end{aligned}$$

Therefore, $\frac{1}{k^2} \geq \Pr(|X-\mu| \geq k\sigma)$ and we are done. □