# Mathematical statistics

February $25^{th}$, 2018

Lecture 6: Statistics and sampling distribution (cont.)

# Topics

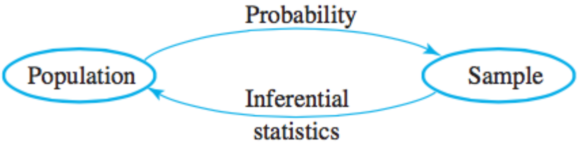
| | |
|---|---|
| **Week 1** | Probability reviews |
| **Week 2** | **Chapter 6: Statistics and Sampling Distributions** |
| **Week 4** | Chapter 7: Point Estimation |
| **Week 7** | Chapter 8: Confidence Intervals |
| **Week 10** | Chapter 9: Test of Hypothesis |
| **Week 14** | Regression |

# Overview

- 6.1 Statistics and their distributions
- 6.2 The distribution of the sample mean
- 6.3 The distribution of a linear combination

Order $6.1 \rightarrow 6.3 \rightarrow 6.2$

# Random sample




**Definition**

The random variables $X_1, X_2, ..., X_n$ are said to form a (simple) random sample of size $n$ if

1. the $X_i$'s are independent random variables
2. every $X_i$ has the same probability distribution

# Questions for this chapter

Given a random sample $X_1, X_2, \ldots, X_n$, and

$$T = a_1X_1 + a_2X_2 + \ldots + a_nX_n$$

- If we **know** the distribution of $X_i$'s, can we obtain the distribution of $T$?
- If we **don't know** the distribution of $X_i$'s, can we still obtain/approximate the distribution of $T$?

# 6.1: Summary

1. If the distribution and the statistic $T$ is simple, try to construct the pmf of the statistic (as in Example 1)
2. If the probability density function $f_X(x)$ of $X$'s is known, the
   - try to represent/compute the cumulative distribution (cdf) of $T$
   $$\mathbb{P}[T \leq t]$$
   - take the derivative of the function (with respect to $t$ )

# 6.2: Linear combination of normal random variables

**Theorem**

*Let $X_1, X_2, \ldots, X_n$ be independent normal random variables (with possibly different means and/or variances). Then*

$$T = a_1X_1 + a_2X_2 + \ldots + a_nX_n$$

*also follows the normal distribution.*

What are the mean and the standard deviation of $T$?

- $E(T) = a_1E(X_1) + a_2E(X_2) + \ldots + a_nE(X_n)$
- $\sigma_T^2 = a_1^2\sigma_{X_1}^2 + a_2^2\sigma_{X_2}^2 + \ldots + a_n^2\sigma_{X_n}^2$

# Moment generating function

**Definition**

The moment generating function (mgf) of a continuous random variable $X$ is

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$$

Reading: 3.4 and 4.2

# Moment generating function

**Property**

*Two distributions have the same pdf if and only if they have the same moment generating function*

# Moment generating function

| Distribution | Moment-generating function $M_X(t)$ |
|---|---|
| Bernoulli $P(X=1)=p$ | $1-p+pe^t$ |
| Geometric $(1-p)^{k-1}\,p$ | $\dfrac{pe^t}{1-(1-p)e^t}$ $\forall t < -\ln(1-p)$ |
| Binomial B($n$, $p$) | $\left(1-p+pe^t\right)^n$ |
| Poisson Pois($\lambda$) | $e^{\lambda(e^t-1)}$ |
| Uniform (continuous) U($a$, $b$) | $\dfrac{e^{tb}-e^{ta}}{t(b-a)}$ |
| Uniform (discrete) U($a$, $b$) | $\dfrac{e^{at}-e^{(b+1)t}}{(b-a+1)(1-e^t)}$ |
| Normal $N(\mu, \sigma^2)$ | $e^{t\mu+\frac{1}{2}\sigma^2t^2}$ |
| Chi-squared $\chi^2_k$ | $(1-2t)^{-\frac{k}{2}}$ |
| Gamma $\Gamma(k, \theta)$ | $(1-t\theta)^{-k}; \forall t < \dfrac{1}{\theta}$ |
| Exponential Exp($\lambda$) | $\left(1-t\lambda^{-1}\right)^{-1},\ (t<\lambda)$ |
| [illegible] | [illegible] |

# Moment generating function

**Definition**

Let $X_1, X_2$ be a 2 independent random variables and $T = X_1 + X_2$, then

$$M_T(t) = M_{X_1}(t)M_{X_2}(t)$$

Hint:

$$M_T(t) = E(e^{tT}) = E(e^{t(X_1+X_2)}) = E(e^{tX_1} \cdot e^{tX_2})$$

# Example 3

## Problem

*Given that the mgf of a Poisson variables with mean* $\lambda$ *is*

$$e^{\lambda(e^t - 1)}$$

*Suppose* $X$ *and* $Y$ *are independent Poisson random variables, where* $X$ *has mean* $a$ *and* $Y$ *has mean* $b$*. Show that* $T = X + Y$ *also follows the Poisson distribution.*

# Example 4

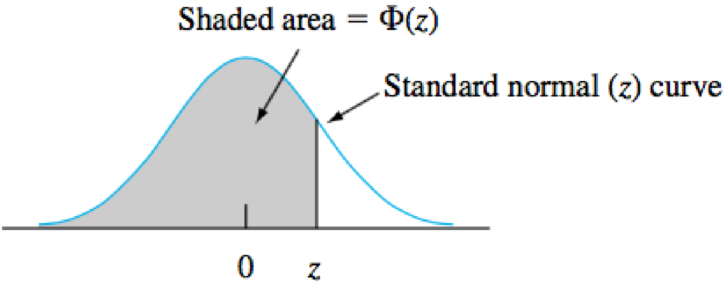
**Problem**

*Given that the mgf of a normal random variables with mean $\mu$ and variance $\sigma^2$ is*

$$e^{\mu t + \frac{\sigma^2}{2} t^2}$$

*Suppose $X$ and $Y$ are independent normal random variables. Show that $T = X + Y$ also follows the normal distribution.*

# $\Phi(z)$




$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} f(y)\, dy$$

# $\Phi(z)$

**Table A.3** Standard Normal Curve Areas (*cont.*) $\Phi(z) = P(Z \leq z)$

| $z$ | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | .5000 | .5040 | .5080 | .5120 | .5160 | .5199 | .5239 | .5279 | .5319 | .5359 |
| 0.1 | .5398 | .5438 | .5478 | .5517 | .5557 | .5596 | .5636 | .5675 | .5714 | .5753 |
| 0.2 | .5793 | .5832 | .5871 | .5910 | .5948 | .5987 | .6026 | .6064 | .6103 | .6141 |
| 0.3 | .6179 | .6217 | .6255 | .6293 | .6331 | .6368 | .6406 | .6443 | .6480 | .6517 |
| 0.4 | .6554 | .6591 | .6628 | .6664 | .6700 | .6736 | .6772 | .6808 | .6844 | .6879 |
| 0.5 | .6915 | .6950 | .6985 | .7019 | .7054 | .7088 | .7123 | .7157 | .7190 | .7224 |
| 0.6 | .7257 | .7291 | .7324 | .7357 | .7389 | .7422 | .7454 | .7486 | .7517 | .7549 |
| 0.7 | .7580 | .7611 | .7642 | .7673 | .7704 | .7734 | .7764 | .7794 | .7823 | .7852 |
| 0.8 | .7881 | .7910 | .7939 | .7967 | .7995 | .8023 | .8051 | .8078 | .8106 | .8133 |
| 0.9 | .8159 | .8186 | .8212 | .8238 | .8264 | .8289 | .8315 | .8340 | .8365 | .8389 |
| 1.0 | .8413 | .8438 | .8461 | .8485 | .8508 | .8531 | .8554 | .8577 | .8599 | .8621 |
| 1.1 | .8643 | .8665 | .8686 | .8708 | .8729 | .8749 | .8770 | .8790 | .8810 | .8830 |
| 1.2 | .8849 | .8869 | .8888 | .8907 | .8925 | .8944 | .8962 | .8980 | .8997 | .9015 |
| 1.3 | .9032 | .9049 | .9066 | .9082 | .9099 | .9115 | .9131 | .9147 | .9162 | .9177 |
| 1.4 | .9192 | .9207 | .9222 | .9236 | .9251 | .9265 | .9278 | .9292 | .9306 | .9319 |
| 1.5 | .9332 | .9345 | .9357 | .9370 | .9382 | .9394 | .9406 | .9418 | .9429 | .9441 |
| 1.6 | .9452 | .9463 | .9474 | .9484 | .9495 | .9505 | .9515 | .9525 | .9535 | .9545 |
| 1.7 | .9554 | .9564 | .9573 | .9582 | .9591 | .9599 | .9608 | .9616 | .9625 | .9633 |
| 1.8 | .9641 | .9649 | .9656 | .9664 | .9671 | .9678 | .9686 | .9693 | .9699 | .9706 |
| 1.9 | .9713 | .9719 | .9726 | .9732 | .9738 | .9744 | .9750 | .9756 | .9761 | .9767 |

# Example 1

## Problem

*Let* $X_1, X_2, \ldots, X_{16}$ *be a random sample from* $\mathcal{N}(1, 4)$ *(that is, normal distribution with mean* $\mu = 1$ *and standard deviation* $\sigma = 2$*).*
*Let* $\bar{X}$ *be the sample mean*

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_{16}}{16}$$

- *What is the distribution of* $\bar{X}$*?*
- *Compute* $P[\bar{X} \leq 1.82]$

# Example 1*

**Problem**

*Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathcal{N}(\mu, \sigma^2)$ (that is, normal distribution with mean $\mu$ and standard deviation $\sigma$).*
*Let $\bar{X}$ be the sample mean*

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

*What is the distribution of $\bar{X}$?*