Economic Statistics - Midterm 2.2 Cheat Sheet

by lunarorbit via cheatography.com/216484/cs/47304/

Probability basics (definitions & rules)

Probability Basics

What probability actually means:

• Probability is a number that tells you how likely something is to happen.

• It's always between 0 and 1:

‣ 0 = impossible

‣ 1 = guaranteed

‣ 0.5 = 50% chance

You can think of probability as the **long-run frequency** of something happening if you repeated it a bunch of times.

Example: if you flip a fair coin 1,000 times, about half the flips will be heads → P(Heads)=0.5.

Key probability symbols

• Event: any outcome or collection of outcomes you're interested in.

Example: "Rolling an even number" on a die → that's the event A = {2, 4, 6}.

• Sample space (S): all possible outcomes.

Example: rolling a die → S = {1, 2, 3, 4, 5, 6}.

• P(A): "The probability that event A happens."

Example: P(Even)=3/6=0.5.

Types of events

• Disjoint / mutually exclusive: can't happen at the same time.

Example: "Rolling a 3" and "Rolling a 4."

• Independent: one happening doesn't change the chance of the other.

Example: Coin toss and rolling a die — they don't affect each other.

Example: "At least one"

"If there's a 0.004 chance a test gives a false positive, what's the probability that at least one out of 200 tests is a false positive?"

This uses the **complement rule** — it's easier to find the chance that none are false positives, then subtract from 1.

P(At least one)=1−P(None)

Each test has a 0.996 chance of being fine → P(None)=0.996^200

So P(At least one)=1−0.996^200

Random Variables (RVs)

What they are

A random variable is just a number that represents the outcome of something random.

Example:

Toss a coin twice.

X = number of heads.

Possible X values: 0, 1, or 2.

We can describe all the possible values of X and how likely they are. That list is called a **probability distribution.**

Probability distribution

A table or formula showing:

Every possible value of X

The probability of each value

Must satisfy:

1. All probabilities are between 0 and 1.

2. They add up to 1.

Mean (Expected Value)

The **expected value (E[X])** or **mean (μ_x)** tells you the average outcome in the *long run.*

E[X]=∑(x×P(X=x))

Example (from table above):

E[X]=0(0.25)+1(0.50)+2(0.25)=1

→ On average, 1 head per 2 flips.

Variance & Standard Deviation

Var(X)=∑(x−μ)² P(X=x) / SD(X)=√Var(X)

Variance = average of squared distances from the mean.

Standard deviation = average distance.

If SD is small → values are close to the mean.

Useful shortcuts

• *E[aX+b]=aE[X]+b*

(Multiply and add constants outside the expectation.)

• *Var(aX+b)=a² Var(X)*

(Adding doesn't affect spread, multiplying stretches it.)

• *E[X+Y]=E[X]+E[Y]*

• If X and Y independent: *Var(X+Y)=Var(X)+Var(Y)*

• If correlated: *Var(X+Y) = Var(X)+Var(Y)+2ρσxσy*

(ρ = correlation between *X* and *Y*)

Example: Lottery

You buy a $1 ticket that pays $500 if you win (probability = 1/1000). If you lose, you get $0.

E[X]=(499)(0.001)+(−1)(0.999)=−0.5

You lose 50¢ on average each play → expected loss.

Sampling & Sampling Distributions

Population vs Sample

• **Population:** the entire group you care about (all students in a school).

• **Sample:** the smaller group you actually measure (30 students).

We use samples to **estimate** the truth about populations.

Economic Statistics - Midterm 2.2 Cheat Sheet

by lunarorbit via cheatography.com/216484/cs/47304/

Types of samples

• **Simple Random Sample (SRS):** every individual has an equal chance of being chosen.
• **Stratified sample:** population divided into groups (strata) → random sample from each.
• **Cluster / multistage:** randomly pick clusters, then pick within them.
⚠☐ **Voluntary response or convenience samples are biased!** (not random → results unreliable).

Sampling distribution

If you take **many random samples** and compute a statistic (like a mean or proportion) for each sample, the **distribution of those statistics** is the sampling distribution.
We study its:
• **Center:** average (should equal population value if unbiased)
• **Spread:** how much sample results vary
• **Shape:** often becomes bell-shaped for large samples

Unbiased estimator

A statistic is **unbiased** if its sampling distribution's center equals the true population parameter.
$E[\bar{X}]=\mu$ and $E[\hat{p}]=p$
Unbiased means, "on average, it hits the right answer."

How sample size affects variability

Bigger sample → smaller variability (less spread).
$SD(\bar{X})=\sigma/\sqrt{n}$
As n increases, denominator gets bigger → SD gets smaller.

Law of Large Numbers & Central Limit Theorem

Law of Large Numbers

When you take more and more samples, the sample mean X̄ will get closer and closer to the population mean μ.
Example: the average of 10 coin flips might not be 0.5, but the average of 10,000 flips almost definitely will be close to 0.5.

Central Limit Theorem (CLT)

This is *super important* for exams.

Even if the original population is **not Normal**, when you take **a large enough sample**, the **distribution of sample means** will look **Normal (bell-shaped)**.
$\bar{X} \sim N(\mu, \sigma/\sqrt{n})$
Meaning:
Centered at μ (the true mean)
Spread = $\sigma/\sqrt{n}$

Why it matters

CLT lets you use **z-scores** and Normal tables to find probabilities for sample means.
*z = sample mean - μ / σ/√n)
Then use the z-table (or calculator) to find probabilities.

Example (CLT in action)

A machine fills cereal boxes.
μ = 500g, σ = 10g.
Sample 25 boxes.
$SD(\bar{X})$ = 10/√25 = 2
$P(\bar{X} < 496) = P(Z <$ 496-500/2) = $P(Z <$ -2) = 0.0228
→ 2.28% chance the sample average weight is below 496g.

Probability rules

RULE	MEANING	FORMULA	EX
1. Range	Probabilities go from 0 to 1	0≤P(A)≤1	A 70% chan not 1.3
2. Whole Sample	All outcomes together have probability 1	P(S)=1	Something
3. Addition (Disjoint)	If A and B can't both happen, add them	P(A or B)=P(A)+P(B)	Rolling a 3 → 1/6
4. Complement	Opposite events add to 1	P(A^c)=1−P(A)	If 30% rai
5. General Addition	If A and B can overlap	P(A or B)=P(A)+P(B) −P(A and B)	Subtract o doubl

Probability Distribution Chart

X	P(X)
0	0.25
1	0.50
2	0.25

Quick strategy for exam problems

STEP	WHAT TO CHECK
1.	Identify if question is about sin (probability), random variables (expectation/variance), or sam (means/proportions).
2.	Draw or imagine the sample sp the possible outcomes?
3.	If you see words like "at least," use probability rules (addition
4.	If you see "average of samples of successes," use sampling fo
5.	Convert to z-score when findir of sample means or proportion
6.	Watch for independence: if eve independent, you can't multiply probabilities.

Formula Summary

CONCEPT	FORMULA	MEANING
Probability of A	P(A)	Likelihood tha happens
Complement	P(A^c)=1−P(A)	Probability th happen
Addition (disjoint)	P(A or B)=P(A)+P(B)	If events can'
Addition (general)	P(A or B)=P(A)+P(B)−P(A and B)	Subtract over
Independence	P(A and B)=P(A)P(B)	Only if A, B in
Expected value	E[X]=∑xP(X = x)	Long-run ave
Variance	Var(X)=∑(x−μ)² P(X=x)	Spread aroun
Standard deviation	SD(X)= √Var(X)	Average dista
Linear rules	E[aX+b]=aE[X]+b, Var(aX-+b)=a² Var(X)	Adjust for sca
Sample mean	E[X̄]= μ, SD(X̄)=σ/√n	Mean & sprea means
Sample proportion	E[p̂]= p, SD(p̂) = √p(1-p)/n	Mean & sprea proportions
Z-score (sampling)	z = sample mean - μ / sd	Converts to S for probability