BAN 602: Quantitative Fundamentals

Lecture 2: Probability and Probability Distributions

Agenda

1. Big Picture of Chapter 4: Probability Basics
2. Random Experiment, Sample Space (and My Critiques)
3. Joint Prob. Table, Conditional Probability, Independent Events, and Bayes Theorem
4. Big Picture of Chapter 5 and 6
5. Basics of Random Variable and (Joint) Probability Distribution

Big Picture of Chapter 4: Probability

<table>
<tr><th>Random Experiments, Counting Rules, and Assigning Probabilities</th><th>Events and Basic Relationships of Probabilities</th><th>Conditional probability (on something)</th></tr>
<tr><td>

- **Random experiments and sample space**
- Counting rules
  - Multi-step experiments (tree diagram)
  - combinations
  - Permutations
- Assigning probabilities
  - Classical (equally likely outcomes)
  - Relative frequency
  - Subjective methods
  - Rules: (1) prob. of each outcome is between 0 and 1; (2) sum of the prob. of all outcomes is 1.

</td><td>

- Event (collection of sample points or experiment outcomes)
- Complement of an event
- Union and intersection (joint) of events and their probabilities
  - Addition law
  - Mutually exclusive events
  - collectively exhaustive events

</td><td>

- **Joint probability table: joint probability and marginal probability**
- **Conditional probability**
  - **Independent events**
  - **Multiplication law**
- **Bayes' Theorem**
  - **prior probability**
  - **posterior probabilities**
  - **total probability (It's nothing but marginal prob)**
  - probability tree or Bayes' theorem

</td></tr>
</table>

Random Experiment and My Critiques

- Random experiment: a process that generates (1) **well-defined experimental outcomes (sample points)**. On any single (2) **repetition** or trial, the outcome that occurs is determined (3) **completely by chance**.
- Sample space is the collection of all possible experiment outcomes. Experiment outcomes are mutually exclusive and collectively exhaustive.
- (4) Black swan: low-probability but high-impact

Frequency (Two-Way) Table or Cross-Tabulation

	Promoted	Not Promoted	Total
M	60	60	120
F	30	50	80
Total	90	110	200

Relative Frequency Table (**Joint Probability Table**)

	Promoted (P)	Not Promoted (NP)	marginal
M	$Pr(M \cap P)=0.3$	$Pr(M \cap NP)=0.3$	$Pr(M)=0.6$
F	$Pr(F \cap P)=0.15$	$Pr(F \cap NP)=0.25$	$Pr(F)=0.4$
marginal	$Pr(P)=0.45$	$Pr(NP)=0.55$	1

- M and F (P and NP) are complement; mutually exclusive; and collectively exhaustive events.
- Random experiment: randomly select an officer for gender. M and F are the 2 possible experiment outcomes. Random variable X (gender) = 1 (M) or 0 (F).
- Random experiment: randomly select an officer for promotion status. P and NP are the 2 possible experiment outcomes. Random variable Y = 1 (P) or 0 (NP).

Joint Probability Distribution

	Y = 1	Y = 0	marginal
X = 1	Pr(X=1, Y=1)=0.3	0.3	Pr(X=1)=0.6
X = 0	0.15	0.25	Pr(X=0)=0.4
marginal	Pr(Y=1)=0.45	Pr(Y=0)=0.55	1

Conditional Probability Definition and Multiplication Rule

Definition: $Pr(A|B) = Pr(AB)/Pr(B)$ or $Pr(A \cap B)/Pr(B)$

Multiplication Rule: $Pr(AB) = Pr(A|B)P(B)$ or $Pr(AB) = Pr(B|A)P(A)$.

Conditional Probability Explanation with Example

- Probability that a male officer is promoted = Pr(P|M) (There is ~~30%~~ 50% chance a male officer is promoted.)

  Pr(P|M) = # of promoted male officers (60)/# of male officers (120) = 50%

  = frequency of promoted male officers/frequency of male officers

  Pr(P|M) = (freq. of promoted male officers/200)/(freq. of male officers/200)

  = relative freq. of promoted male officers/related freq. of male officers

  = Pr(M∩P)/Pr(M) = 0.3/0.6 = 50% or 0.5.

(Statistically) Independent Events

- Are M and P independent?
- More intuitive definition of independent events:
  - If $Pr(A|B) = Pr(A)$, then events A and B are independent. Alternatively, if $Pr(B|A) = Pr(B)$, then events A and B are independent.
- More convenient definition of independent events:
  - If $Pr(AB) = Pr(A)Pr(B)$, then events A and B are independent.
-

(Statistically) Independent Random Variables

- Are X (gender) and Y (promotion status) independent?
- Let $x_1, \ldots, x_m$ be the m possible values of X; Let $y_1, \ldots, y_n$ be the n possible values of Y. Random variables X and Y are independent if and only if $Pr(x_i y_j) = Pr(x_i)Pr(y_j)$ for any pair of i and j.
  - $Pr(x_i y_j)$ is the same as $Pr(X=x_i, Y=y_j)$.
  - $Pr(x_i)$ is the same as $Pr(X=x_i)$; $Pr(y_j)$ same as $Pr(Y=y_j)$

Example

Random experiment: randomly draw a card from a standard deck of 52 cards.

- Event E_1= Ace; E_2= spade. Are E_1 and E_2 independent? Why or why not?

- Random variable X = rank of a card (Ace, King, …, 2); Y = suit of a card (hearts, diamonds, spades, and clubs). Are X and Y independent? Why or why not?

Answer

	Y=1 (S)	Y=2 (H)	Y=3 (D)	Y=4 (C)	marginal prob
X=13 (King)	1/52	1/52	1/52	1/52	1/13
X=12 (Queen)	1/52	1/52	1/52	1/52	1/13
X=11 (Jack)	1/52	1/52	1/52	1/52	1/13
10	1/52	1/52	1/52	1/52	1/13
9	1/52	1/52	1/52	1/52	1/13
8	1/52	1/52	1/52	1/52	1/13
7	1/52	1/52	1/52	1/52	1/13
6	1/52	1/52	1/52	1/52	1/13
5	1/52	1/52	1/52	1/52	1/13
4	1/52	1/52	1/52	1/52	1/13
3	1/52	1/52	1/52	1/52	1/13
2	1/52	1/52	1/52	1/52	1/13
X=1 (Ace)	1/52	1/52	1/52	1/52	1/13
	1/4	1/4	1/4	1/4	1

Interpretation of Independence

- If we know the rank of a card (X), does it help us know better about the suit of this card?
- Prior probability: Pr(Y=y); Posterior probability: Pr(Y=y|X=x)
  - Posterior prob. depends on the information available. Here, Pr(X=x|Y=y) is NOT posterior prob.
- If X and Y are independent, then the information on X does not help us know better about Y. Knowing the rank of a card does not help us know better about the suit.
- Mathematically, Pr(Y=y) = Pr(Y=y|X=x) or Pr(X=x, Y=y) = Pr(X=x)Pr(Y=y).
- Conceptually, prior probability = posterior probability.

Hidden Variable - Frequency

Male	Promoted	Not Promoted	
age<=40	50	30	80
age>40	10	30	40
	60	60	120

Female	Promoted	Not Promoted	
age<=40	12	6	18
age>40	18	44	62
	30	50	80

Conditional Probability Table 1

Define A = (age<=40)

Male	Promoted	Not Promoted	
age<=40	$Pr(P\|M\cap A)=0.625$	0.375	1
age>40	0.25	$Pr(NP\|M\cap A^c)=0.75$	1

Female	Promoted	Not Promoted	
age<=40	0.67	$Pr(NP\|F\cap A)=0.33$	1
age>40	$Pr(P\|F\cap A^c)=0.29$	0.71	1

Conditional Probability Table 2

Male	Promoted	Not Promoted
age<=40	5/6	1/2
age>40	1/6	1/2
	1	1

Female	Promoted	Not Promoted
age<=40	0.4	0.12
age>40	0.6	0.88
	1	1

Application of Joint Probability Table, Conditional Probability and Bayes Theorem

Typically in studying the relationship between two random variables:

A. marginal probabilities about one random variable are available, often called prior probabilities;
B. The conditional probabilities in which the known random variable serves as the condition are available;
C. Joint probabilities are usually not available, but can be derived from A & B;
D. Marginal probabilities about the other random variable, sometimes called total probabilities, are not available, but can be derived from A, B and C.
E. The conditional probabilities in which the unknown random variable serves as the condition, often called posterior probabilities, are not available, but can be derived from A, B, C &D.

Example of JPT & Bayes Theorem (Quiz 2)

Medical studies have shown that 10 out of 100 adults have heart disease. When a person with heart disease is given an EKG test, a 0.9 probability exists that the test will be positive. When a person without heart disease is given an EKG test, a 0.95 probability exists that the test will be negative. Suppose that a person arrives at an emergency room complaining of chest pains. An EKG test is given to this person.

If the test is positive, what is the probability that this person actually has heart disease?

If the test is negative, what is the probability that this person actually does NOT have heart disease?

Example of JPT & Bayes Theorem (Quiz 2)

1. What are the two random variables whose relationship we try to study? What are their sample spaces? Name them appropriately.
2. Information available to us
   a. Which marginal probabilities (about which random variable) do we know? Write them down with previously defined notations. Recall that these probabilities are also called prior prob.
   b. Which conditional probabilities do we know? Write them down with previously defined notations. Note that the known random variable must be the condition.
3. Questions to answer with joint probability table
   a. All the joint probabilities
   b. The other marginal probabilities (about the unknown random variable), sometimes called total probabilities.
   c. The other conditional probabilities with the unknown random variable being the condition, sometimes called posterior probabilities.

Big Picture of Ch. 5: Basics of RV & Prob. Dist.

Random Variable and Its Prob. Dist.	Bivariate (Multivariate) Dist.
• Discrete random variable • Continuous random variable • Question: How to develop prob. dist.? • **Expected value or Expectation** $E(aX+b) = aE(X) + b$ • **Variance and standard deviation** $Var(X) = E(X^2) – (E(X))^2$ $Var(aX+b) = a^2*Var(X)$ • **Prob. mass function vs prob. density function (pmf vs pdf)** • **Cumulative dist. function (cdf)** • **Inverse cdf**	• **Joint probability distribution** • **Covariance** $Cov(X, Y)$ or $\sigma_{XY} = E(XY) – E(X)*E(Y)$ $Cov(X, Y) = cov(Y, X)$ $Cov(a_1X+b_1, a_2Y+b_2) = a_1a_2Cov(X, Y)$ • **correlation** $corr(X, Y) = \sigma_{XY}/(\sigma_X \sigma_Y)$ $corr(aX+b, cY+d) = corr(X, Y)$ • **independence of two RVs X and Y** • $E(X+Y) = E(X) + E(Y)$ $E(XY) = E(X)*E(Y)$ if X & Y are independent $Var(aX+bY) = a^2*Var(X) + b^2*Var(Y)$ $+2ab*\sigma_{XY}$

Big Picture of Ch. 5+6: Discrete & Continuous Prob. Dist.

Discrete Probability Distributions	Continuous Probability Distributions
• **Bernoulli/binary distribution** **f(x)/F(x)** **E(X), Var(X)**, etc. • Binomial distribution f(x)/F(x) E(X), Var(X), etc. • Poisson distribution f(x)/F(x) E(X), Var(X), etc. • Hypergeometric distribution f(x)/F(x) E(X), Var(X), etc.	• Uniform distribution f(x)/F(x) E(X), Var(X), etc. • Normal/Gaussian distribution f(x)/F(x) E(X)=μ, Var(X)=σ^2, sd = σ, etc. • Exponential distribution f(x)/F(x) E(X), Var(X), etc.

Random Variable and Its (Probability) Distributions

A numeric description of the outcome of a random experiment.

How to describe a random variable?

	Discrete RV	Continuous RV
$f(x)$	$f(x) = Pr(X = x)$ prob. mass func.	$f(x) = F'(x)$ prob. density func.
$F(x)$	$F(x) = Pr(X <= x)$	$F(x) = Pr(X <= x)$
$F^{-1}(p)$	$F^{-1}(p) = x$ such that $Pr(X <= x) = p$	$F^{-1}(p) = x$ such that $Pr(X <= x) = p$
$E(X)$		
σ^2 & σ		

Bernoulli/Binary Distribution (Indicator Function)

f(x)

F(x)

E(X)

Var(X)

Example (Part of Quiz 3)

The joint probability distribution of two random variables X and Y are given below:

	X = -1	X = 0	X = 1	
Y = 0	0	1/2	0	1/2
Y = 1	1/4	0	1/4	1/2
	1/4	1/2	1/4	1

Answer the following questions:

1. pmf of X? cdf of X?
2. Compute E(X), E(Y), E(X^2), E(Y^2), E(XY), Var(X), Var(Y), cov(X, Y), corr(X, Y).
3. X and Y independent? Why or why not?
4. Consider (A): X and Y are (linearly) uncorrelated; (B) cov(X, Y) = 0; (C) corr(X, Y) = 0; (D) X and Y are independent.

Probability Distribution Example (If Time Allows)

For borrower with good credit scores, the mean debt for revolving and installment accounts is $15,015. Assume the standard deviation is $3540 and that debt amount is normally distributed.

What is the probability that the debt for a borrower with good credit is less than $10K?

```
pnorm(10000, 15015, 3540) = 0.104837
```

Probability Distribution Example (If Time Allows)

For borrower with good credit scores, the mean debt for revolving and installment accounts is $15,015. Assume the standard deviation is $3540 and that debt amount is normally distributed.

What is the probability that the debt for a borrower with good credit is more than $18K?

```
1 - pnorm(18000, 15015, 3540) = 0.19955
```

Probability Distribution Example (If Time Allows)

For borrower with good credit scores, the mean debt for revolving and installment accounts is $15,015. Assume the standard deviation is $3540 and that debt amount is normally distributed.

What is the probability that the debt for a borrower with good credit is between $12K and $18K?

pnorm(18000, 15015, 3540) - norm.cdf(12000, 15015, 3540) = 0.387162

Probability Distribution Example (If Time Allows)

For borrower with good credit scores, the mean debt for revolving and installment accounts is $15,015. Assume the standard deviation is $3540 and that debt amount is normally distributed.

If a borrower with good credit scores is in the 85th percentile, what is his debt?

```
qnorm(0.85, 15015, 3540) = 18683.97
```