🍏

Statistics

IMC Correlation Type Questions

Leetcode for Quant Finance

👨‍💻 Morgan Stanley

Awesome Medium Stats Summary

Coursera Computational Finance Summary

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/b5eb59c4-3ae7-4268-8be3-758f77f6cb8b/QuantitativePrimer.pdf

Quantitative Primer Interview Questions Github

Citadel Co-relation One Test

Generating Functions in Simple Words

Stats Learning Course

👨‍💻 Squarepoint Capital

Aman (BRO) DS MS Intern test

Yifan Zhao told that Econometric sort-off questions are asked in hedge funds

💪🏼 Econometrics

🚢 Bayesian Statistics-2

Leetcode for Statistics - HireLLgraph - Kunal told

Naveen Gupta

got asked lot of Expectation Q

MIT- Fundamentals of Statistics Edx

MIT- Statistics for Applications

Problem	Definition	Consequences	Detection	Solution
High multicollinearity	Two or more independent variables in a regression model exhibit a close linear relationship.	Large standard errors and insignificant *t*-statistics Coefficient estimates sensitive to minor changes in model specification Nonsensical coefficient signs and magnitudes	Pairwise correlation coefficients Variance inflation factor (VIF)	1. Collect additional data. 2. Re-specify the model. 3. Drop redundant variables.
Heteroskedasticity	The variance of the error term changes in response to a change in the value of the independent variables.	Inefficient coefficient estimates Biased standard errors Unreliable hypothesis tests	Park test Goldfeld-Quandt test Breusch-Pagan test White test	1. Weighted least squares (WLS) 2. Robust standard errors
Autocorrelation	An identifiable relationship (positive or negative) exists between the values of the error in one period and the values of the error in another period.	Inefficient coefficient estimates Biased standard errors Unreliable hypothesis tests	Geary or runs test Durbin-Watson test Breusch-Godfrey test	1. Cochrane-Orcutt transformation 2. Prais-Winsten transformation 3. Newey-West robust standard errors

Citadel Co-relation One Test

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/adeddfe6-2ee8-47c4-b262-187456982b60/Quant_Interview_Prep.pdf

Stats for Dummies Cheat Sheet

Stats Topics to Cover

Linda PIMCO Class has some good questions

Manish shared these concise notes

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/4fc83fb2-b83a-41b5-a5a6-081e1dc64f39/SVD.pdf https://s3-us-west-2.amazonaws.com/secure.notion-static.com/1eeee6f6-6c73-4925-bec2-38e8c887f231/Symmetric_Matrices.pdf

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/8c5dda1c-da31-4532-b43d-78c6d3728287/Statistics_Notes.pdf

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/d8237e1e-bc07-45e5-87ea-1a7bd1946eb4/Econometrics_wooldridge_summary_(1).pdf

TERMS TO WRITE ABOUT

- What is AIC and BIC, how do they differ?
- Give an Intuitive Feel of P-test, T-test in OLS?
- What is Adjusted-R-squared?
- Kalman Filter
- GARCH and Arch Volatility Models
- Co-integration
- VAR and VECM
- Maximum Likelihood Estimator
- Best Linear Predictor (BLP)
- Markov Chain Monte Carlo Simulation
- Cholesky Decomposition
- QR Decomposition
- Black Scholes Derivation

- Mean Efficient Frontier
- CAPM
- Confidence Intervals calculation in OLS and what are they in general?

Diagonalizability of a matrix?

Multicollinearity

Notice this peculiar differentiation of Matrix for getting Beta Formula

Regression Theory → Linear, Logistic, MLE Assumptions

Simple Linear Regression

Logistic Regression & Probit Regression

Significance Test → Linear vs Logistic

Bayesians vs Frequentists?

Questions
- Maven Questions
- SIG Questions
- OLS Green Book
  - Gamble → 100 fair coins → If more than 60 heads → +10 else nothing → Should you play this for 1 dollar?
  - 2 pieces of wood, minimise the variance to measure them if only 2 measures are allowed
- Questions asked Personally
  - If there are N stocks with different distributions, then how will the portfolio of stocks behave?
  - How do you find confidence intervals? What is the definition of confidence intervals?
- Apeksha Citadel
  - When do and when don't AIC, BIC work?
  - If you add a feature to your feature set and it results in an increase in the AIC value, would you include that feature in your feature set?
  - Now let's say you have a data set and there is nothing flagged in the correlation matrix of the data set. How would you choose features then?
  - What's the drawback of using PCA?
  - How would you choose features from the feature set which are highly correlated?
  - Code the calculation of pi using two uniformly distributed random variables
  - Ridge regression is used to handle data sets with multicollinearity issue, how do you think ridge regression would perform in this case where the features are perfectly correlated?
  - If you have data sets with m columns and you duplicate 1 column and add p-1 such duplicate features to your set, would OLS work? How would you handle it?
- Other MFE Interviews
  - How to convert uniform distribution to exponential distribution?
  - Execution strategies, Is VWAP the most optimal price?
  - I flip a coin 100 times and get 40 heads. Is the coin fair? Construct a statistical test that would help answer this question.
  - If the correlation between two random variables is 0, does this mean they are independent?

- You sample 2 points from the population: 1 and -1. Can you calculate the variance?
- Explain linear regression and its assumptions. What is regularization? How would you expect the estimates of betas to differ for OLS, L1, and L2?

Manish FAQ List

- Impact of Multicollinearity in data
- Impact of imbalanced data and missing values on performance of regression
- How to compute square root to a certain digit?
- L1 and L2 regularization
- Properties of Correlation matrices
- Calculate var(x) given that the data points distribute uniformly on 3D sphere
- How would you measure the risk of an equity portfolio
- Linear Regression, Regularization, feature selection
- How to convert a uniform random variable to a normal random variable. You have r red balls, w white balls in a bag. If you keep drawing balls out of the bag until the bag now only contains balls of a single color (ie you run out of a color) what is the probability you run out of white balls first? (in terms of r and w)
- What is correlation? What is covariance? Made a graph that the correlation is equal to 1 and -1.
- What datasets would you utilize to predict quarterly sales for company X?
- VWAP
- Expected flips to see 2 heads
- information theory, what is entropy, significance of bit
- What's the difference of regress y on x and regress x on y?
- You are holding two eggs in a 100-story building. If an egg is thrown out of the window, it will not break if the floor number is less than X, and it will always break if the floor number is equal to or greater than X. What strategy would you use to determine X with the minimum number of drops in a worst case scenario?
- State and prove Central Limit theorem
- Blindfolded. In a room with 8 coins. What is the minimum number of flips required to guarantee that one of the permutations had all coins with the same side?
- Chance that a student passes the test is 10%. What is the chance that out of 400 students AT LEAST 50 pass the test? Check the closest answer.

GSA Capital Test

Classical ML Questions

- GTS→ How does random forest work and explain bagging?
- Do you know in-sample error and out-sample error ?
- some covariance matrix estimation machine learning questions
- Why SVM outperforms other classifiers?
- What is the difference between GNB and logistic regression? Which one would you opt under different circumstances?

- Pros and cons of Naive Bayes classifier, Random forest, Linear regression, Neural Networks, Logic regression

OLS Questions

- What happens when measurement error is in X vs when it is in Y?
- What is PACF intuitively?
- How would Beta change if i have errors in measurement of X and Y variables?
- Give the definition of R^2 in the context of linear regression. Can it be negative? Prove that it can. How would you interpret this situation?
- You regress Y on X and then X on Y. Can you say anything about how betas in those regressions are related?
- What remains same and what changes on changing Input?
- Different variations asked in 👨‍💻 GTS

Expectation Questions

- Both X and Y are ~ N(0,1). What can you say about the distribution of X + Y?
- Kunal Blackstone question → E(X|X>Y) where X and Y are standard normal with correlation p
- Finding $E(X \mid X>Y)$ when X,Y are i.i.d $U(0,1)$
- E[X| X>Y] for independent X, Y ~ N(0,1)
- If $X,Y \sim N(0,1)$, what is $P(X>3Y \mid Y>0)$?
- What is probability $P(X<x|X<Y)$
- PIMCO Question → X, Y → 2 Standard normal variables, how will you calculate value of X-Y absolute value being less than $sqrt(2)$
- X,Y are iid standard normal. what $P(Y>3X)$
- Derive E(X4) where X ∼ Normal(0, σ2)
- Given log X ~ N(0,1). Compute the expectation of X.
- If Rho_{xy} = 0, and x and y are normal, are they independent? Give an example where they are dependent and it is still zero
- Covariance Calculation of Uniform and Exponential Random Variable
- If X is $N(\mu, \sigma)$, then what is Distribution and expectation of X^2
- Find Correlation of X, Y variables that are inside a Circle of unit radius
- Let X and Y be two independent Uniform(0,1) random variables. Let also Z=max(X,Y) and W=min(X,Y). Find Cov(Z,W)
- More Solved Problems

Bivariate Distribution and Sum of Normal Distributions

- Conditional Distribution of Y given X=x, where X and Y are jointly normal?
- Solved Problems

C Crib sheet

Print this sheet and keep it with you during phone interviews, but make sure you are able to reproduce it wholly from memory at some point in your life. Actual familiarity with the concepts is better than rote, but a combination of the two is required for good interview performance.

Miscellaneous

ARMA(p, q):

$$Y_t = \alpha_0 + \underbrace{\sum_{i=1}^{p} \alpha_i Y_{t-i}}_{\text{Autoregressive}} + \overbrace{\sum_{i=1}^{q} \beta_i \varepsilon_{t-i}}^{\text{Moving average}} + \varepsilon_t$$

GARCH(p, q):

$$Y_t = \sigma_t Z_t \quad \text{where } Z_t \sim N(0,1)$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2$$

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$$

$${}_nC_r = \binom{n}{k} = \frac{n!}{(n-k)!k!}$$

$$\sum_{k=0}^{n-1} ar^k = a\left(\frac{1-r^n}{1-r}\right)$$
$$\sum_{k=0}^{\infty} ar^k = \frac{a}{1-r} \text{ for } |r| < 1$$

Regression assumptions

- The covariates are observed without error (weak exogeneity)
- Linearity
- Constant variance (homoscedasticity)
- Independence of errors
- No (or very little) multicollinearity
- Statistical distribution, if required

Linear regression

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \beta_0}{\text{stdev}(\tilde{\beta})} \sim t(v = n - p)$$

Simple linear regression

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$
$$\hat{\beta} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\text{Cov}(x,y)}{\text{Var}(x)} = \rho_{xy}\frac{s_y}{s_x}$$

$$R^2 = \rho_{yx}^2$$

Logistic regression

$$W_{\hat{\beta}} = \frac{(\hat{\beta} - \beta_0)^2}{\text{stdev}(\tilde{\beta})} \sim \mathcal{X}^2$$

Calculus

The following are important for most brainteasers:

- Product rule of differentiation:

$$\frac{d}{dx} g(x)f(x) = g'(x)f(x) + g(x)f'(x)$$

- Integration by parts:

$$\int_a^b u\,dv = uv\Big|_{x=a}^{x=b} - \int_a^b v\,du$$

- Integration and differentiation rules involving e^x and $\ln(x)$.

TERMS TO WRITE ABOUT

What is AIC and BIC, how do they differ?

BIC is more conservative, and is the preferred method these days, example in the cancelling roots case in Lab 3, We saw that BIC had lower scores for AR-1 than ARMA-2,1

(Look at Apeksha's question of Citadel)

Give an Intuitive Feel of P-test, T-test in OLS?

What is Adjusted-R-squared?

```
r_squared_adj = r_squared - (1 - r_squared)*((k - 1)/(Nobs - k))
Nobs = y.shape[0] # Number of Sample Points
k = X.shape[1] # Number of features
```

Kalman Filter

GARCH and Arch Volatility Models

Co-integration

VAR and VECM

Maximum Likelihood Estimator

Best Linear Predictor (BLP)

Markov Chain Monte Carlo Simulation

Cholesky Decomposition

QR Decomposition

Black Scholes Derivation

Why it doesn't have a drift term?

Mean Efficient Frontier

CAPM

Confidence Intervals calculation in OLS and what are they in general?

How it is calculated is in Lab 4 230E section

Diagonalizability of a matrix?

Multicollinearity

If Multicollinearity is present:

1. It often leads to overfitting
2. t-tests can no longer be trusted
3. All Features must be included, irrespective of what the features say

- *No multicollinearity:* if the covariates are correlated, they can still be used in the regression, but numerical problems might occur depending on how the fitting algorithms invert the matrices involved. The t-tests that the regression produces can no longer be trusted. All the covariates must be included regardless of what their significance tests say. A big problem with multicollinearity, however, is over-fitting. Depending on how bad the situation is, the parameter values might have huge uncertainties around them, and if you fit the model using new data their values might change significantly. I suggest reading the Wikipedia article on multicollinearity, as it contains useful information:
  `https://en.wikipedia.org/wiki/Multicollinearity`
  Multicollinearity is a favourite topic of discussion for interviewers, and they usually have strong opinions about how it should be handled. The model's intended use will determine how sensitive it is to ignoring the error distribution. In many cases, fitting a line using least-squares estimation is equivalent to assuming errors have a normal distribution. If the real distribution has heavier tails, like the t-distribution, how risky will it make decisions based on your outputs? One way to address this is to use a technique like robust-regression. Another way is to think about the dynamics behind the problem and which distribution would be best suited to model them—as opposed to just fitting a curve through a set of points.

Notice this peculiar differentiation of Matrix for getting Beta Formula

$$\varepsilon = Y - X\beta$$

and you have the following total sum of squared errors,

$$\begin{aligned} S(\beta) &= \varepsilon^T \varepsilon \\ &= (Y - X\beta)^T (Y - X\beta). \end{aligned}$$

For the least-squared-error estimate, you want to find the β that minimises the sum of squared errors. Take the derivative and set it equal to zero:

$$\begin{aligned} S(\beta) &= Y^T Y - \beta^T X^T Y - Y^T X\beta + \beta^T X^T X\beta \\ &= Y^T Y - 2Y^T X\beta + \beta^T X^T X\beta && \text{(result is } n \times 1) \\ \frac{\partial}{\partial\beta} S(\beta) &= -2X^T Y + 2X^T X\beta && \text{(result is } p \times 1) \\ \frac{\partial^2}{\partial\beta\partial\beta^T} S(\beta) &= 2X^T X && \text{(result is } p \times p). \end{aligned}$$

The easiest way to convince yourself of the matrix derivatives is to work with 2×2 matrices. Next, set the first derivative equal to zero to find a critical point of the function,

$$\begin{aligned} \frac{\partial}{\partial\beta} S(\beta) &= 0 \\ 2X^T X\beta - 2X^T Y &= 0 \\ X^T X\beta &= X^T Y \\ \beta &= (X^T X)^{-1} X^T Y \end{aligned}$$

giving the required result. This is a minimum since the second derivative is a positive definite matrix.

Regression Theory→ Linear, Logistic, MLE Assumptions

Given nicely in section 3 of Quantitative Primer Book

> In order to user Likelihood in MLE, we make assumption on the error, that individual errors follow normal distribution, so error vector is multivariate normal distribution

> We finally reach the same expression for Beta in MLE as was in Linear regression

Simple Linear Regression

R-Squared meaning → Proportion of variance in Y explained by variance in X

3.2 Simple linear regression

The Simple linear regression is the special case of the linear regression with only one covariate

$$y = \alpha + x\beta,$$

75

which is just a straight line fit. Interviewers like this model for its aesthetically pleasing theoretical properties. A few of them are described here, beginning with parameter estimation. For n pairs of (x_i, y_i),

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

minimise the sum of squared errors,

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \alpha + \beta x_i)^2,$$

as before, but without matrices. The least-squares estimators are:

$$\begin{aligned} \hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x} \\ \hat{\beta} &= \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \\ &= \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)} && (18) \\ &= \rho_{xy} \frac{s_y}{s_x}. && (19) \end{aligned}$$

Use

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

where Var and Cov are the variance and covariance, respectively. Likewise, ρ_{xy}, s_y, and s_x are the sample correlation coefficient and sample standard deviations. The final two lines of the equation, (18) and (19), are of utmost importance. Learn them by heart, as they make many questions about simple linear regression much easier. For example, "when is the slope of a straight-line fit through a set of points (x_i, y_i) equal to the correlation between x and y?"

The final important result is that, for a simple linear regression, the R^2 value (which measures the proportion of variance in y_i explained by the variance in x_i) is equal to the sample correlation coefficient squared,

$$R^2 = \rho_{yx}^2.$$

Beta = Cov(x,y)/Variance(x) → This is the main stuff

Correlation Coefficient(x,y) = **Cov(x,y) / (std(x) std(y))**

So, Beta = Correlation Coefficient(x,y) . **std(x) . std(y)/ std(x) . std(x)**

Beta = Correlation Coefficient(x,y) . **std(y)/ std(x)**

For Simple Linear Regression R-Squared Value = Correlation Coefficient(x,y)^2

Logistic Regression & Probit Regression

R-squared doesnt exist for Logistic

3.3 Logistic regression

Both the logistic regression and linear regression are GLMs, as introduced by McCullagh (1984), but users might spend years using regressions without ever encountering this term. The techniques have different names in different fields, a phenomenon that has been made worse by the recent "rediscovery" of linear regression and logistic regression in Machine Learning. It would be prudent to know some specialist theory and important results about each of the three regressions discussed in this section, using the terminology that pertains to finance.

The assumptions for the logistic regression from the perspective of a GLM are identical to those of the linear regression, except that you are assuming linearity on the logit scale. Let the

variate of interest y be a binary variable taking values of 0 and 1, then

$$\begin{aligned} y &\sim \text{Bernoulli}(p) \\ \text{logit}(p) &= X\beta \quad \text{(logit link function)} \\ &\text{or} \\ p &= \text{logistic}(X\beta). \end{aligned}$$

The only difference between the logistic and the probit regression is the link function. For the probit regression you have

$$\begin{aligned} y &\sim \text{Bernoulli}(p) \\ \text{Probit}(p) &= X\beta \quad \text{(Probit link function)} \\ &\text{or} \\ p &= \Phi(X\beta), \end{aligned}$$

where Φ is the cumulative distribution function (CDF) of the standard normal distribution, and $\text{Probit}(x)$ is the quantile function (or inverse CDF).[12] The R-square measure as defined for the linear regression doesn't exist for the logistic regression. There are a few measures that act as pseudo R-squared, but I won't discuss them here—see the Wikipedia page for details.[13]

Significance Test → Linear vs Logistic

3.1.3 Significance tests

In the ordinary linear regression, the t-statistic is used to test the significance of each parameter,

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \beta_0}{\mathrm{stdev}(\tilde{\beta})}$$

where $\mathrm{stdev}(\tilde{\beta})$ is the standard deviation of the estimator $\tilde{\beta}$. For the classical linear regression with homoscedasticity and normal errors, the sampling distribution of $t_{\hat{\beta}}$ is the student's t-distribution with $(n-p)$ degrees of freedom. Here, β_0 is a hypothesised value to test $\hat{\beta}$ against, usually set to 0.

Linear Regression

When multi-collinearity is present, not much emphasis is placed on interpretation of t-test because std. dev will be high and Beta_cap would be correlated, t-statistic is not equal to Wald-Statistic. Wald Statistic is used in case of Logistic regression.

3.3.1 Significance tests

Where the linear regression uses the t-statistic for significance testing, the logistic regression has the Wald-statistic. Asymptotically, it has a chi-squared distribution:

$$W_{\hat{\beta}} = \frac{(\hat{\beta} - \beta_0)^2}{\mathrm{stdev}(\tilde{\beta})} \sim \mathcal{X}^2$$

where β_0 is a hypothesised value, as with the linear regression.

Logistic Regression

Bayesians vs Frequentists?

a frequentist would need to maximise the likelihood us- ing a numerical technique, whereas a Bayesian would evaluate the posterior distribution numerically, or sample from it using MCMC.

Questions

Maven Questions

Decent Probability Question

Another Decent Probability Question

SIG Questions

Biased Coin, throws to equal head and tails

An empty tank that has a hole in it

Max of two die rolls expectation

[Interview experience] SIG Quant electric gluten | Just look at dry goods

Direct

Interview experience | Mathematics | Metalworking | sig

herbyu10 2017-2-13 16:31:55 | Just look at the author | Try Job

This building: 0% (0) 0% (0)
Global: 100% (82) 0% (0)

2017 (January-March) **Full-time PhD in Metalworking @ SIG** -Online Haitou- **Technical Communication** | **Fail/Rej** | Job-hopping

Contribute a gluten, I hope it will be helpful to the interview partners below

SIG, a company doing prop trading in Philadelphia, has a very different interview style from the trading firm in Chicago, New York. It's all statistical probability mathematics. Unfortunately, I have returned to the teacher what I learned in college.

1. Three normal distributions X, Y, Z, corr(X, Y) = corr(Y, Z) = corr(X, Z) = 0.5, find the probability of X + Y> Z+2
2. Some properties of the Correlation matrix: semidefinite, if A^2 = I, what special properties will there be
3. The price of a stock today is 100 yuan. Tomorrow this stock will either become 50 or 200. There is also a casino where you can bet on the rise or fall of stocks. There is no limit to the amount of bet. The odds are 1:1. Ask you how to make a strategy, how to make money? In fact, it's just to get a hedge arbitray, long stock short gambling
4. Given a [0, 1] distribution, how to generate a point (x, y, z) uniformly distributed on the surface of the sphere

OLS Green Book

Gamble → 100 fair coins → If more than 60 heads→ +10 else nothing → Should you play this for 1 dollar?

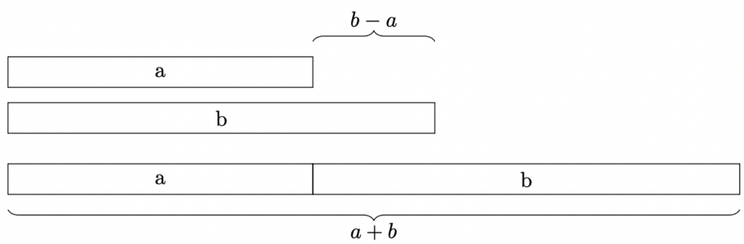
Can be done by converting it into normal distribution

Mean = 50 Variance = 100 x 0.5 x 0.5 = 25

2 pieces of wood, minimise the variance to measure them if only 2 measures are allowed

Measuring a+b and a-b will help reduce variance

Using £2, you've measured each one with an error of σ^2, but you can do better by measuring the sum of the two sticks, $s = a + b$, and the difference between the two $d = a - b$:



Solve for a and b in terms of s and d, and see what the mean and variance of the estimates are.

$$\begin{aligned} a &= \frac{s+d}{2} \\ \mathrm{E}(a) &= \mathrm{E}\left(\frac{s+d}{2}\right) = a \\ \mathrm{Var}(a) &= \mathrm{Var}\left(\frac{s+d}{2}\right) \\ &= \frac{1}{4}(\mathrm{Var}(s) + \mathrm{Var}(d)) \\ &= \frac{1}{2}\sigma^2 \\ &\text{Likewise, for } b \\ b &= \frac{s-d}{2} \\ \mathrm{E}(b) &= \mathrm{E}\left(\frac{s-d}{2}\right) = b \\ \mathrm{Var}(b) &= \mathrm{Var}\left(\frac{s-d}{2}\right) \\ &= \frac{1}{2}\sigma^2 \end{aligned}$$

Using £2 and the new strategy, you've measured each stick with an error of $\sigma^2/2$. What happened here? As if by magic, the variance has been reduced by half. The trick is that you used the tools to get two measurements of a and b each. Each additional measurement will reduce the variance of the estimator.

Questions asked Personally

Moody Analytics

If there are N stocks with different distributions, then how will the portfolio of stocks behave?

Use CLT and everything will behave normally!

How do you find confidence intervals? What is the definition of confidence intervals?

hypothesis testing and CI are basically inverse of hypothesis testing

Apeksha Citadel

When do and when don’t AIC, BIC work?

If you add a feature to your feature set and it results in an increase in the AIC value, would you include that feature in your feature set?

Now let's say you have a data set and there is nothing flagged in the correlation matrix of the data set. How would you choose features then?

What's the drawback of using PCA?

How would you choose features from the feature set which are highly correlated?

Code the calculation of pi using two uniformly distributed random variables

Ridge regression is used to handle data sets with multicollinearity issue, how do you think ridge regression would perform in this case where the features are perfectly correlated?

If you have data sets with m columns and you duplicate 1 column and add p-1 such duplicate features to your set, would OLS work? How would you handle it?

Other MFE Interviews

How to convert uniform distribution to exponential distribution?

Execution strategies, Is VWAP the most optimal price?

I flip a coin 100 times and get 40 heads. Is the coin fair? Construct a statistical test that would help answer this question.

Stack Exchange

If the correlation between two random variables is 0, does this mean they are independent?

You sample 2 points from the population: 1 and -1. Can you calculate the variance?

Explain linear regression and its assumptions. What is regularization? How would you expect the estimates of betas to differ for OLS, L1, and L2?

Regression assumptions

- The covariates are observed without error (weak exogeneity)
- Linearity
- Constant variance (homoscedasticity)
- Independence of errors
- No (or very little) multicollinearity
- Statistical distribution, if required

Manish FAQ List

Impact of Multicollinearity in data

Impact of imbalanced data and missing values on performance of regression

How to compute square root to a certain digit?

Can just do binary search or use newton's method of iteration : use eqn x-squared -3 = 0 as f(x)

L1 and L2 regularization

Properties of Correlation matrices

Calculate var(x) given that the data points distribute uniformly on 3D sphere

How would you measure the risk of an equity portfolio

Linear Regression, Regularization, feature selection

How to convert a uniform random variable to a normal random variable. You have r red balls, w white balls in a bag. If you keep drawing balls out of the bag until the bag now only contains balls of a single color (ie you run out of a color) what is the probability you run out of white balls first? (in terms of r and w)

What is correlation? What is covariance? Made a graph that the correlation is equal to 1 and -1.

What datasets would you utilize to predict quarterly sales for company X?

VWAP

Expected flips to see 2 heads

information theory, what is entropy, significance of bit

What's the difference of regress y on x and regress x on y?

You are holding two eggs in a 100-story building. If an egg is thrown out of the window, it will not break if the floor number is less than X, and it will always break if the floor number is equal to or greater than X. What strategy would you use to determine X with the minimum number of drops in a worst case scenario?

This has the solution explained nicely

The n-egg problem

Problem

Given n eggs and t tries, how many floors can we explore?

Solution

Let's start by dropping one egg: If it breaks, we'll explore the lower $d_{n-1}(t-1)$ floors; if it survives, we'll explore the upper $d_n(t-1)$ floors. Overall, the total distance we can explore is:
$d_n(t) = 1 + d_{n-1}(t-1) + d_n(t-1)$
$d_n(0) = 0$

We've already solved the one- and two-egg problems:
$d_1(t) = t$ for $t \geq 0$
$d_2(t) = \frac{t(t+1)}{2}$ for $t \geq 0$

And, taken together, these statements give us the solution to *every* n-egg problem. In particular, here is the solution to the three-egg problem:

State and prove Central Limit theorem

Blindfolded. In a room with 8 coins. What is the minimum number of flips required to guarantee that one of the permutations had all coins with the same side?

Chance that a student passes the test is 10%. What is the chance that out of 400 students AT LEAST 50 pass the test? Check the closest answer.

GSA Capital Test

Recorded, can see in 20Jul21 recordings

Classical ML Questions

GTS→ How does random forest work and explain bagging?

Do you know in-sample error and out-sample error ?

some covariance matrix estimation machine learning questions

Why SVM outperforms other classifiers?

What is the difference between GNB and logistic regression? Which one would you opt under different circumstances?

Pros and cons of Naive Bayes classifier, Random forest, Linear regression, Neural Networks, Logic regression

OLS Questions

What happens when measurement error is in X vs when it is in Y?

THIS IS AN AWESOME LINK

Thus (important conclusion), measurement error in an independent variable will tend to bias its estimated slope coefficient towards zero in OLS.

What is PACF intuitively?

1. It is the amount of correlation with each lag that is not accounted for by more recent lags
2. Its like we remove the auto-correlation for lesser lag terms and then see how much these other earlier lags weren't able to explain
3. We can calculate this by running regression while controlling for other lags
    a. I think what this means is that just do an OLS with all the lags before it once
    b. then just try to regress residuals with this specific lag term and not the coefficient to report it as PACF for this lag

How would Beta change if i have errors in measurement of X and Y variables?

What if Only X variable? What if Only Y variable?

Beta = covariance / variance

Give the definition of R^2 in the context of linear regression. Can it be negative? Prove that it can. How would you interpret this situation?

It can be negative, it just denotes that fit is more poor than it would have been if a straight horizontal line was being fit

as before, but without matrices. The least-squares estimators are:

$$\begin{aligned}
\hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x} \\
\hat{\beta} &= \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
&= \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)} \qquad (18) \\
&= \rho_{xy}\frac{s_y}{s_x}. \qquad (19)
\end{aligned}$$

Use

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

where Var and Cov are the variance and covariance, respectively. Likewise, ρ_{xy}, s_y, and s_x are the sample correlation coefficient and sample standard deviations. The final two lines of the equation, (18) and (19), are of utmost importance. Learn them by heart, as they make many questions about simple linear regression much easier. For example, "when is the slope of a straight-line fit through a set of points (x_i, y_i) equal to the correlation between x and y?"

The final important result is that, for a simple linear regression, the R^2 value (which measures the proportion of variance in y_i explained by the variance in x_i) is equal to the sample correlation coefficient squared,

$$R^2 = \rho_{yx}^2.$$

You regress Y on X and then X on Y. Can you say anything about how betas in those regressions are related?

This is a good place for checking it out

Simple linear regression

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)} = \rho_{xy}\frac{s_y}{s_x}$$

$$R^2 = \rho_{yx}^2$$

<u>What remains same and what changes on changing Input?</u>

Different variations asked in 👨‍💻 <u>GTS</u>

Expectation Questions

Both X and Y are ~ N(0,1). What can you say about the distribution of X + Y?

This involves Joint Normality

<u>This Stack overflow answer better explains it</u>

(8) The moment generating function corresponding to the normal probability density function $N(x; \mu, \sigma^2)$ is the function $M_x(t) = \exp\{\mu t + \sigma^2 t^2/2\}$.

The notable characteristic of this function is that it is in the form of an exponential. This immediately implies that the sum of two independently distributed Normal random variables is itself a normally distributed random variable. Thus:

(9) Let $x \sim N(\mu_x, \sigma_x^2)$ and $y \sim N(\mu_y, \sigma_y^2)$ be two independently distributed normal variables. Then their sum is also a normally distributed random variable: $x + y = z \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$.

To prove this we need only invoke the result that, in the case of independence, the moment generating function of the sum is the product of the moment generating functions of its elements. This enables us to write

$$\begin{aligned} M_z(t) &= M_x(t)M_y(t) = \exp\left\{\mu_x t + \tfrac{1}{2}\sigma_x^2 t^2\right\} \exp\left\{\mu_y t + \tfrac{1}{2}\sigma_y^2 t^2\right\} \\ &= \exp\left\{(\mu_x + \mu_y)t + \tfrac{1}{2}(\sigma_x^2 + \sigma_y^2)t^2\right\} \end{aligned} \tag{10}$$

which is recognised as the moment generating function of a normal distribution.

Kunal Blackstone question → E(X|X>Y) where X and Y are standard normal with correlation p

Finding $E(X \mid X>Y)$ when X,Y are i.i.d $U(0,1)$

E[X| X>Y] for independent X, Y ~ N(0,1)

If $X,Y \sim N(0,1)$, what is $P(X>3Y \mid Y>0)$?

What is probability $P(X<x|X<Y)$

PIMCO Question → X, Y → 2 Standard normal variables, how will you calculate value of X-Y absolute value being less than $sqrt(2)$

This X-Y is not necessarily normal, It needs to be jointly normal for X-Y to be normal

Most simple case is when X and Y are independent, in that case, X-Y will still have mean as zero, variance will be 2, so → 66% within 1 std. dev.

What about case when they are not independent, if correlation is rho, then we need to find the distribution and see accordingly

X,Y are iid standard normal. what P(Y>3X)

Apeksha was asked in Citadel

9

Yes, it's 1/2, a simple way to solve it is using normal Rv rules, i.e. $P(Y > 3X) = P(Y - 3X > 0)$, and $Z = Y - 3X$ is normal RV with mean 0 and variance 10. Since the mean is 0, independent of the variance, $P(Z > 0)$ is $1/2$, since the normal curve is symmetric around 0.

Your logic is not correct. We can't say that given the variables are positive, $P(Y > 3X)$ is 0, and vice versa for both negative case.

Edit based on request:

$$E[Z] = E[Y - 3X] = E[Y] - 3E[X] = 0$$

$$\operatorname{var}(Z) = \operatorname{var}(Y - 3X) = \operatorname{var}(Y) + (-3)^2 \operatorname{var}(X) = 10$$

Derive E(X4) where X ∼ Normal(0, σ2)

to arrive at the right answer, but when I reached (1) the interviewer said "there, now we have something familiar" (referring to the second moment) and went on to the next question. A better approach might have been to use the Moment Generating Function of the normal distribution,

$$M(t) = e^{\mu t} e^{\frac{1}{2}\sigma^2 t^2}.$$

Then you can calculate any moment using

$$E\left(X^n\right) = M_X^{(n)}(0) = \left.\frac{d^n M_X(t)}{dt^n}\right|_{t=0}$$

(8) The moment generating function corresponding to the normal probability density function $N(x; \mu, \sigma^2)$ is the function $M_x(t) = \exp\{\mu t + \sigma^2 t^2/2\}$.

where $M_X^{(n)}(t)$ is the nth derivative of the function $M(t)$. In the current case where $\mu = 0$,

$$\begin{aligned} M_X(t) &= e^{\frac{1}{2}\sigma^2 t^2} \\ M_X^{(1)}(t) &= \sigma^2 t e^{\frac{\sigma^2 t^2}{2}} \\ M_X^{(2)}(t) &= \sigma^2 \left(\sigma^2 t^2 + 1\right) e^{\frac{\sigma^2 t^2}{2}} \\ M_X^{(3)}(t) &= \sigma^4 t \left(\sigma^2 t^2 + 3\right) e^{\frac{\sigma^2 t^2}{2}} \\ M_X^{(4)}(t) &= \sigma^4 \left(\sigma^4 t^4 + 6\sigma^2 t^2 + 3\right) e^{\frac{\sigma^2 t^2}{2}} \\ M_X^{(4)}(0) &= 3\sigma^4. \end{aligned}$$

Still tedious, but perhaps less so than the integration.[6] Of course, you could just learn the moments of the normal distribution by heart, but I don't think that's what interviewers want—they want to see some mathematics. In my case, the interviewer also asked me what the fourth moment meant. I didn't understand, but they probably wanted me to mention the kurtosis, which is a measure of the "tailedness" of a distribution. The fourth moment is used to measure kurtosis, the third moment is related to the skewness, the second moment to the variance, and the first moment is the mean.

4th Moment is kurtosis and measures tailedness

Given log X ~ N(0,1). Compute the expectation of X.

Can find proof here below

Expected value

The expected value of a log-normal random variable X is

$$\mathrm{E}[X] = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$$

Proof

Variance

The variance of a log-normal random variable X is

$$\mathrm{Var}[X] = \exp\left(2\mu + 2\sigma^2\right) - \exp\left(2\mu + \sigma^2\right)$$

Proof

Higher moments

The n-th moment of a log-normal random variable X is

$$\mathrm{E}[X^n] = \exp\left(n\mu + \frac{1}{2}n^2\sigma^2\right)$$

Proof

If Rho_{xy} = 0, and x and y are normal, are they independent? Give an example where they are dependent and it is still zero

No. Correlation only measures linear dependence. They could have a non-linear dependence

Example is below

Answer 11.a): (Question on page 20)

Let Z have the probability mass function

$$p(z) = \begin{cases} 0.5 & \text{if } z = -1 \\ 0.5 & \text{if } z = 1 \end{cases}$$

or equivalently

$$\begin{aligned} P(Z = -1) &= 0.5 \\ P(Z = 1) &= 0.5. \end{aligned}$$

Let $X \sim \text{Normal}(0, 1)$ and $Y = ZX$. Derive the distribution of Y using its cumulative distribution function:

$$\begin{aligned} P(Y < x) &= P(ZX < x | Z = 1)P(Z = 1) + P(ZX < x | Z = -1)P(Z = -1) \\ &= P(X < x)(0.5) + P(-X < x)(0.5) \\ &= (0.5)(P(X < x) + P(X \geq -x)) \\ &= (0.5)(P(X < x) + P(X < x)) \quad \text{symmetry} \\ &= P(X < x) \end{aligned}$$

meaning Y has the same distribution as X. Now, check whether $\rho = 0$,

$$\begin{aligned} \rho &= \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \\ &= \text{Cov}(X, Y) \\ &= \text{Cov}(X, ZX) \\ &= \text{E}(XZX) - \text{E}(X)\,\text{E}(ZX) \\ &= \text{E}(X^2 Z) - \text{E}(X)\,\text{E}(ZX) \\ &= \text{E}(X^2)\,\text{E}(Z) - \text{E}(X)\,\text{E}(Z)\,\text{E}(X) \quad \text{independence} \\ &= 0 \quad \text{since } E(Z) = 0. \end{aligned}$$

You have successfully constructed a situation where there is a clearly defined dependence between Y and X, but they are both distributed $\text{Normal}(0, 1)$ with correlation coefficient $\rho = 0$.

Covariance Calculation of Uniform and Exponential Random Variable

Example 5.32

Suppose $X \sim Uniform(1,2)$, and given $X = x$, Y is exponential with parameter $\lambda = x$. Find $\text{Cov}(X, Y)$.

Solution

We can use $\text{Cov}(X,Y) = EXY - EXEY$. We have $EX = \frac{3}{2}$ and

$$
\begin{aligned}
EY &= E[E[Y|X]] && \big(\text{law of iterated expectations (Equation 5.17)}\big) \\
&= E\left[\frac{1}{X}\right] && \big(\text{since } Y|X \sim Exponential(X)\big) \\
&= \int_1^2 \frac{1}{x} dx \\
&= \ln 2.
\end{aligned}
$$

We also have

$$
\begin{aligned}
EXY &= E[E[XY|X]] && \big(\text{law of iterated expectations}\big) \\
EXY &= E[XE[Y|X]] && \big(\text{since} E[X|X=x] = x\big) \\
&= E\left[X\frac{1}{X}\right] && \big(\text{since } Y|X \sim Exponential(X)\big) \\
&= 1.
\end{aligned}
$$

Thus,

$$\text{Cov}(X,Y) = E[XY] - (EX)(EY) = 1 - \frac{3}{2}\ln 2.$$

If X is $N(\mu, \sigma)$, then what is Distribution and expectation of X^2

It is the famous chi-squared distribution

Use the identity

$$E(X^2) = \text{Var}(X) + [E(X)]^2$$

and you're done.

Since you know that $X \sim N(\mu, \sigma)$, you know the mean and variance of X already, so you know all terms on RHS.

Find Correlation of X, Y variables that are inside a Circle of unit radius

Law of Iterated Expectations comes handy

Example 5.34

Let X and Y be as in Example 5.24 in Section 5.2.3, i.e., suppose that we choose a point (X, Y) uniformly at random in the unit disc

$$D = \{(x, y) | x^2 + y^2 \leq 1\}.$$

Are X and Y uncorrelated?

Solution

We need to check whether $\text{Cov}(X, Y) = 0$. First note that, in Example 5.24 of Section 5.2.3, we found out that X and Y are not independent and in fact, we found that

$$X|Y \sim Uniform(-\sqrt{1 - Y^2}, \sqrt{1 - Y^2}).$$

Now let's find $\text{Cov}(X, Y) = EXY - EXEY$. We have

$$\begin{aligned} EX &= E[E[X|Y]] && \big(\text{law of iterated expectations (Equation 5.17)}\big) \\ &= E[0] = 0 && \big(\text{since } X|Y \sim Uniform(-\sqrt{1 - Y^2}, \sqrt{1 - Y^2})\big). \end{aligned}$$

Also, we have

$$\begin{aligned} E[XY] &= E[E[XY|Y]] && \big(\text{law of iterated expectations (Equation 5.17)}\big) \\ &= E[YE[X|Y]] && \big(\text{Equation 5.6}\big) \\ &= E[Y \cdot 0] = 0. \end{aligned}$$

Thus,

$$\text{Cov}(X, Y) = E[XY] - EXEY = 0.$$

Thus, X and Y are uncorrelated.

Let X and Y be two independent Uniform(0,1) random variables. Let also Z=max(X,Y) and W=min(X,Y). Find Cov(Z,W)

Key is to start and compute CDF and then differentiate to get PDF, finally just use the covariance formula

It is useful to find the distributions of Z and W. To find the CDF of Z, we can write

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= P(\max(X, Y) \leq z) \\ &= P\Big((X \leq z) \text{ and } (Y \leq z)\Big) \\ &= P(X \leq z)P(Y \leq z) && \text{(since } X \text{ and } Y \text{ are independent)} \\ &= F_X(z)F_Y(z). \end{aligned}$$

Thus, we conclude

$$F_Z(z) = \begin{cases} 0 & z < 0 \\ z^2 & 0 \leq z \leq 1 \\ 1 & z > 1 \end{cases}$$

Therefore,

$$f_Z(z) = \begin{cases} 2z & 0 \leq z \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

From this we obtain $EZ = \frac{2}{3}$. Note that we can find EW as follows

$$\begin{aligned} 1 = E[X + Y] &= E[Z + W] \\ &= EZ + EW \\ &= \frac{2}{3} + EW. \end{aligned}$$

Thus, $EW = \frac{1}{3}$. Nevertheless, it is a good exercise to find the CDF and PDF of W, too. To find the CDF of W, we can write

$$\begin{aligned}
F_W(w) &= P(W \leq w) \\
&= P(\min(X, Y) \leq w) \\
&= 1 - P(\min(X, Y) > w) \\
&= 1 - P\Big((X > w) \text{ and } (Y > w)\Big) \\
&= 1 - P(X > w)P(Y > w) && \text{(since } X \text{ and } Y \text{ are independent)} \\
&= 1 - (1 - F_X(w))(1 - F_Y(w)) \\
&= F_X(w) + F_Y(w) - F_X(w)F_Y(w).
\end{aligned}$$

Thus,

$$F_W(w) = \begin{cases} 0 & w < 0 \\ 2w - w^2 & 0 \leq w \leq 1 \\ 1 & w > 1 \end{cases}$$

Therefore,

$$f_W(w) = \begin{cases} 2 - 2w & 0 \leq w \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

From the above PDF we can verify that $EW = \frac{1}{3}$. Now, to find $\mathrm{Cov}(Z, W)$, we can write

$$\begin{aligned}
\mathrm{Cov}(Z, W) &= E[ZW] - EZEW \\
&= E[XY] - EZEW \\
&= E[X]E[Y] - E[Z]E[W] && \text{(since } X \text{ and } Y \text{ are independent)} \\
&= \frac{1}{2}.\frac{1}{2} - \frac{2}{3}.\frac{1}{3} \\
&= \frac{1}{36}.
\end{aligned}$$

Note that $\mathrm{Cov}(Z, W) > 0$ as we expect intuitively.

More Solved Problems

Bivariate Distribution and Sum of Normal Distributions

Sum of two independent normal random variables is also normal, but if they are not independent, their sum is not necessarily normal, it is normal only when they are jointly normal, this is actually the definition of Jointly Normal Distribution

Definition 5.3
Two random variables X and Y are said to be **bivariate normal**, or **jointly normal**, if $aX + bY$ has a normal distribution for all $a, b \in \mathbb{R}$.

SUM OF NORMALLY DIST VARIABLES

BIVARIATE NORMAL DISTRIBUTION

P(Y>3X| Y>0) → X and Y are iid standard normal distributed

We can construct 2 random Variables of correlation rho in this way

Example 5.36
Let Z_1 and Z_2 be two independent $N(0, 1)$ random variables. Define

$$X = Z_1,$$
$$Y = \rho Z_1 + \sqrt{1-\rho^2} Z_2,$$

where ρ is a real number in $(-1, 1)$.

a. Show that X and Y are bivariate normal.
b. Find the joint PDF of X and Y.
c. Find $\rho(X, Y)$.

The rho(X,Y) is actually rho, it can also be calculated like this below

c. To find $\rho(X, Y)$, first note

$$Var(X) = Var(Z_1) = 1,$$
$$Var(Y) = \rho^2 Var(Z_1) + (1 - \rho^2)Var(Z_2) = 1.$$

Therefore,

$$\begin{aligned}
\rho(X, Y) &= \mathrm{Cov}(X, Y) \\
&= \mathrm{Cov}(Z_1, \rho Z_1 + \sqrt{1 - \rho^2} Z_2) \\
&= \rho \mathrm{Cov}(Z_1, Z_1) + \sqrt{1 - \rho^2} \mathrm{Cov}(Z_1, Z_2) \\
&= \rho \cdot 1 + \sqrt{1 - \rho^2} \cdot 0 \\
&= \rho.
\end{aligned}$$

If we have above expression for X and Y, then we can show that X and Y are bivariate normal by a simple argument

First, note that since Z_1 and Z_2 are normal and independent, they are jointly normal, with the joint PDF

$$\begin{aligned}
f_{Z_1 Z_2}(z_1, z_2) &= f_{Z_1}(z_1) f_{Z_2}(z_2) \\
&= \frac{1}{2\pi} \exp\left\{ -\frac{1}{2}\left[z_1^2 + z_2^2\right] \right\}.
\end{aligned}$$

a. We need to show $aX + bY$ is normal for all $a, b \in \mathbb{R}$. We have

$$\begin{aligned}
aX + bY &= aZ_1 + b(\rho Z_1 + \sqrt{1 - \rho^2} Z_2) \\
&= (a + b\rho) Z_1 + b\sqrt{1 - \rho^2} Z_2,
\end{aligned}$$

which is a linear combination of Z_1 and Z_2 and thus it is normal.

Finding the joint distribution of X and Y can be done in this following way →

We basically use inverse transformation to get values of Z_1 and Z_2, then we use joint distribution of Z_1 x Z_2 and some determinant jacobian expression as well as shown

b. We can use the method of transformations (Theorem 5.1) to find the joint PDF of X and Y. The inverse transformation is given by

$$Z_1 = X = h_1(X, Y),$$
$$Z_2 = -\frac{\rho}{\sqrt{1-\rho^2}}X + \frac{1}{\sqrt{1-\rho^2}}Y = h_2(X, Y).$$

We have

$$\begin{aligned} f_{XY}(z_1, z_2) &= f_{Z_1Z_2}(h_1(x, y), h_2(x, y))|J| \\ &= f_{Z_1Z_2}(x, -\frac{\rho}{\sqrt{1-\rho^2}}x + \frac{1}{\sqrt{1-\rho^2}}y)|J|, \end{aligned}$$

where

$$J = \det\begin{bmatrix} \frac{\partial h_1}{\partial x} & \frac{\partial h_1}{\partial y} \\ \\ \frac{\partial h_2}{\partial x} & \frac{\partial h_2}{\partial y} \end{bmatrix} = \det\begin{bmatrix} 1 & 0 \\ \\ -\frac{\rho}{\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} \end{bmatrix} = \frac{1}{\sqrt{1-\rho^2}}.$$

Thus, we conclude that

$$\begin{aligned} f_{XY}(x, y) &= f_{Z_1Z_2}(x, -\frac{\rho}{\sqrt{1-\rho^2}}x + \frac{1}{\sqrt{1-\rho^2}}y)|J| \\ &= \frac{1}{2\pi}\exp\left\{-\frac{1}{2}\left[x^2 + \frac{1}{1-\rho^2}(-\rho x + y)^2\right]\right\} \cdot \frac{1}{\sqrt{1-\rho^2}} \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[x^2 - 2\rho xy + y^2\right]\right\}. \end{aligned}$$

Conditional Distribution of Y given X=x, where X and Y are jointly normal?

Theorem 5.4

Suppose X and Y are jointly normal random variables with parameters μ_X, σ_X^2, μ_Y, σ_Y^2, and ρ. Then, given $X = x$, Y is normally distributed with

$$E[Y|X = x] = \mu_Y + \rho\sigma_Y\frac{x-\mu_X}{\sigma_X},$$
$$Var(Y|X = x) = (1-\rho^2)\sigma_Y^2.$$

Example 5.37

Let X and Y be jointly normal random variables with parameters μ_X, σ_X^2, μ_Y, σ_Y^2, and ρ. Find the conditional distribution of Y given $X = x$.

Solution

One way to solve this problem is by using the joint PDF formula (Equation 5.24). In particular, since $X \sim N(\mu_X, \sigma_X^2)$, we can use

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}.$$

Another way to solve this problem is to use Theorem 5.3. We can write

$$\begin{cases} X &= \sigma_X Z_1 + \mu_X \\ Y &= \sigma_Y(\rho Z_1 + \sqrt{1-\rho^2} Z_2) + \mu_Y \end{cases}$$

Thus, given $X = x$, we have

$$Z_1 = \frac{x - \mu_X}{\sigma_X},$$

and

$$Y = \sigma_Y \rho \frac{x - \mu_X}{\sigma_X} + \sigma_Y \sqrt{1-\rho^2} Z_2 + \mu_Y.$$

Since Z_1 and Z_2 are independent, knowing Z_1 does not provide any information on Z_2. We have shown that given $X = x$, Y is a linear function of Z_2, thus it is normal. In particular

$$\begin{aligned} E[Y|X=x] &= \sigma_Y \rho \frac{x - \mu_X}{\sigma_X} + \sigma_Y \sqrt{1-\rho^2} E[Z_2] + \mu_Y \\ &= \mu_Y + \rho \sigma_Y \frac{x - \mu_X}{\sigma_X}, \\ Var(Y|X=x) &= \sigma_Y^2 (1-\rho^2) Var(Z_2) \\ &= (1-\rho^2)\sigma_Y^2. \end{aligned}$$

We conclude that given $X = x$, Y is normally distributed with mean $\mu_Y + \rho\sigma_Y \frac{x-\mu_X}{\sigma_X}$ and variance $(1-\rho^2)\sigma_Y^2$.

Solved Problems