Lecture 3: Value-at-Risk

Gordon J. Ross

Value at Risk

This week we are going to be discussing **Value-at-Risk**, which is a widely used measure used by banks, pensions funds, etc for assessing the risk associated with a portfolio of financial assets.

It is also used outside finance in situations where it is necessary to quantify **how much an individual or institution stands to lose** over some given period of time

Value at Risk - Introduction

Suppose a company holds a certain number of financial assets such as stocks, bonds, derivatives, etc. This collection of assets is called a **portfolio**.

Every day, the value of the portfolio will change, since the price of the individual assets in the portfolio change frequently. The value of the portfolio can go up or down. If it goes down, the company obviously makes a loss.

The company typically wants to know the answer to questions such as ”what is the probability of the value of the portfolio dropping by more than $10 million on a given day?”. Knowing this is essential to risk management.

Value at Risk - Definition

The rough decision of VaR is ”what is the most I can lose on an investment?”.

Or rather, ”what is the most I can lose on an investment, with some probability p”

It is hence a measure of **risk exposure**

Value at Risk - Definition

For any particular value p, the Value-at-Risk of a given portfolio is the threshold such that the probability of losing more than this threshold on a given day is equal to p.

For example if $p = 0.05$ and the VaR is $\$1m$, this means that there is a 5% probability of losing more than $\$1m$ in a single day.

Similarly if $p = 0.01$ and the VaR is $\$10m$, this means that there is a 1% probability of losing more than $\$10m$ in a single day

Typically we will be given the portfolio, and p, then asked to find the VaR – i.e. for risk management the company will choose some value of p (e.g. $p = 0.01$) then try to find what sum of money it has a 100p% chance of losing in a single day

Value at Risk - Definition

In this course, rather than working with monetary loss, we will instead usually focus on **percentage loss**, since this is typically more informative.

For example if $p = 0.05$ and the VaR is 2%, this means that there is a 5% probability that the value of the portfolio will decrease by more than 2% in a single day.

Similarly if $p = 0.01$ and the VaR is 4%, this means that there is a 1% probability that the value of the portfolio will decrease by more than 4% in a single day

Value at Risk - Estimation

There are many ways of computing the VaR, which have varying degrees of sophistication. We will start with a very simple method (which is widely used in practice - perhaps it should not be!). We will discuss more sophisticated methods as we proceed.

In the simple method, we assume that the percentage change in the portfolio on each day t is a random variable Y_t, **and that the percentage change on each day is independent of the percentage change on all other days**

In other words, If V_t denotes the value of the portfolio on day t, then $V_t = Y_t \times V_{t-1}$ where $Y_t \sim F$ is a random variable drawn from some distribution F. By the above assumption, the Y_t values are **independent and identically distributed**

Value at Risk - Estimation

The VaR will be estimated based on historical data – i.e based on how the portfolio has behaved in the past. Suppose the company has held the portfolio for $n+1$ days. They hence have n observations $Y = (Y_1, \ldots, Y_n)$ where each Y_r denotes the percentage change in the value of the portfolio on day $t+1$

Eg suppose that on the first day the portfolio was worth \$100m, then on the second day it was worth \$99m, then on the third day it was worth \$101m, then on the fourth it was worth \$102m.

The first 3 values of Y_t are then:

$$Y_1 = \frac{99}{100} = 0.99 \quad \text{(a 1\% loss)}$$

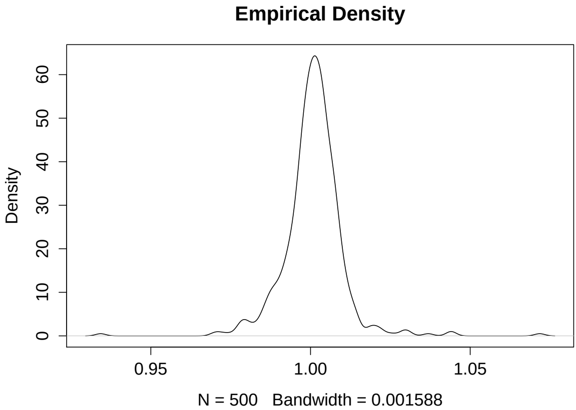
$$Y_2 = \frac{101}{99} = 1.0202 \quad \text{(a 2.02\% gain)}$$

$$Y_3 = \frac{102/}{101} = 1.0099 \quad \text{(a 0.99\% gain)}$$

Values smaller than 1 denote a loss.

Percentage Loss - Empirical Density

For most portfolios, the empirical density/histogram of the previous n Y_t values will look something like this:



Again: values smaller than 1 denote a loss.

Value at Risk - Computation

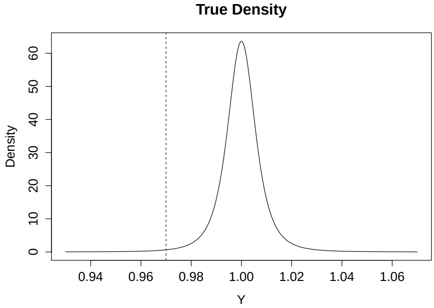
Suppose we know the true density F (where $Y_t \sim F$). Then we can compute the VaR as follows:

Find the $100p^{th}$ quantile of the **true** density of Y_t. Denote this by q. For example if $p = 0.01$ then q is equal to the 1^{st} quantile, i.e.we find q to satisfy $p(Y_t < q) = 0.01$. In other words, q is the point at which the total area under the curve below that point is equal to 0.01

Then q is the VaR.

Percentage Loss - Computation

E.g. if the following is the true density of Y_t and $p = 0.01$...



Then $q = 0.97$. This means that there is a 1% probability that the value of the portfolio will drop by 3% or more on any given day (since $Y_t = 0.97$ denotes a 3% loss).

Value at Risk - Calculation

In the **simplest possilble model** we might assume that the distribution of each Y_t is Gaussian, $Y_t \sim N(\mu, \sigma^2)$ Then, we would estimate μ and σ, and use this to compute q.

We use the historical data $Y_1, Y_2, \ldots, Y_n$ to estimate μ and σ^2.

First suppose we are frequentists. The maximum likelihood estimates are the usual $\hat{\mu} = 1/n \sum_{t=1}^{n} Y_t$, and $\hat{\sigma}^2 = 1/(n-1) \sum_{t=1}^{n} (Y_t - \hat{\mu})^2$

For a given p, a simple estimate of q would be the value satisfying:

$$\int_{\infty}^{q} p(Y_t|\hat{\mu}, \hat{\sigma}^2) dY_t = p$$

This can be found easily in R using the quantile function: qnorm(1-p, muest, sigma2est) where muest and sigma2est are the estimated mean and variance

Value at Risk - Simple Frequentist Example

For example, suppose the empirical mean of the historical Y_t was 1.01, with standard deviation 0.01. This means that on average the portfolio gains 1% in value each day, but the returns have a standard deviation of 1%

If $p = 0.01$ then q is given by:

$$qnorm(0.01, 1.01, 0.01) = 0.9867$$

So there is a 1% chance of the portfolio dropping by 1.32% or more in a given day.

(Note:$1 - 0.9867 = 0.0132$)

Value at Risk - Bayesian Approach

Several problems with this simple frequentist approach:

1. Doesnt take uncertainty about μ and σ^2 into account - we used point estimates (MLE). But these estimates will not be accurate - we will underestimate risk because we are not talking their uncertainty into account. We should really be doing something with confidence intervals - but this is difficult
2. Doesnt incorporate prior information about the future portfolio returns – using historical data is important, but we may also have beliefs about the future which aren't reflected in previous history
3. The usual issues about the difficulty of communicating frequentist statements to non-statisticians, and how they get misinterpreted

We will hence explore a Bayesian approach to VaR analysis instead. In this case we need to perform Bayesian inference for the parameters μ, σ^2 of the Gaussian distribution. As you might expect, this is very general and has applications beyond VaR analysis - essentially anywhere that a Gaussian distribution is used!

Value at Risk - Bayesian Approach

Suppose we are Bayesians. We start with a prior distribution $p(\mu, \sigma^2)$ on the unknown parameters of the Gaussian distribution governing the percentage daily change Y_t.

This prior is chosen to reflect our beliefs about the future portfolio returns. Remember: μ_t is the average return on a given day $E[Y_t]$, and σ^2 is the variance $Var[Y_t]$.

We then update this to get the posterior $p(\mu, \sigma^2 | Y_1, \ldots, Y_n)$ given the historical data (we will discuss how to do this in a few slides).

This posterior then captures all our knowledge about the distribution of Y_t based on both our prior knowledge, and the historical data. We can then obtain q based on this.

Value at Risk - Calculation

So, the first question is: how do we go about computing the posterior $p(\mu, \sigma^2|Y)$?

We saw how to do this in Lecture 1 when the likelihood $p(Y|\theta)$ was Binomial, and the prior $p(\theta)$ was a Beta distribution.

We now need to choose a suitable prior $p(\mu, \sigma^2)$ and update this given the Gaussian likelihood $p(Y|\mu, \sigma^2)$

This is a good time for a general discussion about how to choose prior distributions in Bayesian statistics. For example in Lecture 1, why did we use a Beta prior?

A General Discussion of How to Choose Prior Distribution

Recall Lecture 1...

In Lecture 1 we used Bayesian inference to estimate the probability θ of a biased coin landing heads. Here Y denotes the number of heads obtained in N coin tosses.

We used a Beta prior on θ:

$$p(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}$$

and the likelihood was Binomial:

$$p(Y|\theta) = \binom{N}{Y}\theta^{Y}(1-\theta)^{N-Y}$$

Recall Lecture 1...

To learn about θ from the data, we derived the posterior $p(\theta|Y)$, which by Bayes theorem was:

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{\int p(Y|\theta)p(\theta)d\theta}$$

The integral in the denominator looked difficult...

$$\int p(Y|\theta)p(\theta)d\theta = \int \binom{N}{Y} \frac{\theta^{Y+\alpha-1}(1-\theta)^{N-Y+\beta-1}}{B(\alpha,\beta)} d\theta$$

But we used a trick where we recognised this had the form of a Beta distribution, which allowed us to integrate it. The posterior ended up being Beta, just like the prior!

Recall Lecture 1...

Since the posterior distribution had a standard form (Beta), it was easy to analyse. However in many situations things will not be as simple, and the posterior might end up being an unknown distribution, or one which can't be solved analytically.

The key point here is the integral in the denominator of Bayes theorem: $p(Y) = \int p(Y|\theta)p(\theta)d\theta$. Typically if we can solve this integral analytically then the posterior will be easy to analyse. But in many cases it will be impossible to do this integral (remember: outside of A-Level mathematics classes, "most" integrals cannot be solved analytically!).

In this case the integral must be solved numerically instead. We will discuss this at length in a future lecture. But for now, lets focus on the cases where this integral can be solved analytically.

Note that we solved it here by **recognising that it had the same form as a Beta distribution**. This was not a coincidence!

Choice of Prior Distribution

To make the posterior distribution easy to analyse mathematically, we often choose priors which are **conjugate to the likelihood**. Conjugacy means that the posterior distribution has the same form as the prior distribution - for example a Beta prior with a Beta posterior, or a Gamma prior leading to a Gamma posterior, etc. I.e. we deliberately choose the prior so that the posterior has the same form (Beta in this case).

When using conjugate priors, we can always solve the integral $p(Y) = \int p(Y|\theta)p(\theta)d\theta$ analytically by using the trick we used in Lecture 1 – i.e "recognising that it has the same form as the prior" and must integrate to 1 due to being a probability distribution., This makes the posterior easy to analyse. So, choosing a conjugate prior makes things much simpler. How do we find conjugate priors?

Conjugate Priors

Definiton: If $p(\theta)$ belongs to the same family of probability distributions as $p(\theta|Y)$ then $p(\theta)$ is a conjugate prior for θ

From Bayes theorem, focusing only on the numerator, this will be true if $p(\theta|Y)$ has the same general form as $p(\theta)p(Y|\theta)$. By 'general form' I mean "the parts which depend on θ.

In our example the likelihood was Binomial:

$$p(Y|\theta) = \binom{N}{Y}\theta^Y(1-\theta)^{N-Y}$$

In order for the prior to keep its same form when multiplied by this, the part depending on θ must be proportional to:

$$p(\theta|Y) \propto \theta^r(1-\theta)^s$$

for some constants r and s. But this is just the form of the Beta distribution! So the Beta distribution is the conjugate prior for the Binomial distribution.

Conjugate Priors

In general to find a conjugate prior, either use the argument from the previous slide to derive one, or consult a standard table (wikipedia has a very good list on the ”Conjugate Prior” page)

n.wikipedia.org/wiki/Conjugate_prior   wikipedia conjugate priors

assimilation.

Table of conjugate distributions [edit]

Let *n* denote the number of observations. In all cases below, the data is assumed to consist of *n* points $x_1, \ldots, x_n$ (which will be random vectors in the multivariate cases).

If the likelihood function belongs to the exponential family, then a conjugate prior exists, often also in the exponential family; see Exponential family: Conjugate distributions.

Discrete distributions [edit]

Likelihood	Model parameters	Conjugate prior distribution	Prior hyperparameters	Posterior hyperparameters	Interpretation of hyperparameters[note 1]
Bernoulli	*p* (probability)	Beta	$\alpha,\ \beta$	$\alpha + \sum_{i=1}^n x_i,\ \beta + n - \sum_{i=1}^n x_i$	$\alpha - 1$ successes, $\beta - 1$ failures[note 1]
Binomial	*p* (probability)	Beta	$\alpha,\ \beta$	$\alpha + \sum_{i=1}^n x_i,\ \beta + \sum_{i=1}^n N_i - \sum_{i=1}^n x_i$	$\alpha - 1$ successes, $\beta - 1$ failures[note 1]
Negative Binomial with known failure number *r*	*p* (probability)	Beta	$\alpha,\ \beta$	$\alpha + \sum_{i=1}^n x_i,\ \beta + rn$	$\alpha - 1$ total successes, $\beta - 1$ failures[note 1] (i.e. $\frac{\beta - 1}{r}$ experiments, assuming r stays fixed)
Poisson	λ (rate)	Gamma	$k,\ \theta$	$k + \sum_{i=1}^n x_i,\ \frac{\theta}{n\theta + 1}$	k total occurrences in $1/\theta$ intervals
Poisson	λ (rate)	Gamma	$\alpha,\ \beta$ [note 3]	$\alpha + \sum_{i=1}^n x_i,\ \beta + n$	α total occurrences in β intervals
Categorical	***p*** (probability vector), *k* (number of categories, i.e. size of ***p***)	Dirichlet	$\boldsymbol{\alpha}$	$\boldsymbol{\alpha} + (c_1, \ldots, c_k)$, where c_i is the number of observations in category *i*	$\alpha_i - 1$ occurrences of category i[note 1]

An Example

Suppose that in a particular region of the world, N earthquakes have occurred over the last 2000 years. Their occurrence times are $t_1, t_2, \ldots, t_N$. Under the most simple model of seismicity, these earthquakes are assumed to follow a Poisson process, in which case the time-between-events $\tau_i = t_i - t_{i-1}$ follow an Exponential distribution with parameter λ.

For the purpose of predicting the occurrence of future earthquakes, we wish to learn about λ. I.e. given the independent and identically distribtuted observations $\tau_1, \ldots, \tau_{N-1}$ where $\tau_i \sim \mathit{Exponential}(\lambda)$, we wish to infer λ.

As before, we start with a prior distribution which represents our knowledge about λ before analysing the data. Lets try to find a conjugate prior.

An Example - Conjugate Prior

Since the data is Exponential, the likelihood is:

$$p(\lambda|Y) = \lambda e^{-\lambda Y}$$

For the prior distribution to keep its same form after being multiplied by the likelihood, it must be proportional to:

$$p(\lambda) \propto \lambda^r e^{-s\lambda}$$

for some r and s. If we consult list of probability distributions, we find that distribution with this form is the Gamma distribution:

$$p(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}$$

So this is the conjugate prior for the Exponential distribution.

Note: if you do not recognise these forms (or have never seen the Gamma distribution before) then don't worry – again, the conjugate priors for most common probability distributions can just be looked up in a table.

An Example

As always, we choose the parameters of the prior α and β to reflect our prior beliefs about λ. These will be based on either seismological theory, or evidence from other similar earthquake regions.

Your task: Have a go at Question 5 on Exercise Sheet 1, which involves finishing off this example to compute the (Gamma) posterior distribution corresponding to a particular choice of α and β in the prior. The methodology is exactly the same as we used to compute the Beta posterior in Lecture 1.

Back to Value at Risk

Bayesian Inference for the Gaussian Distribution - Known Variance

We now discuss how to perform Bayesian inference for parameters of the Gaussian distribution which describes the daily percentage changes in the portfolio value. The historical returns are $Y = (Y_1, \ldots, Y_n)$ where each $Y_t \sim N(\mu, \sigma^2)$.

Lets start with the simplest case. Suppose that the variance σ^2 is **known**. In this case we only need to estimate μ (which, remember, denotes the average daily percentage change in value)

It can be shown (see lecture supplementary material on moodle) that the conjugate prior in this case is also Gaussian: $p(\mu) = N(\mu_0, \sigma_0^2)$ where μ_0 and σ_0^2 control the shape of the prior, and hence reflect prior beliefs about μ_0.

In other words, we represent our prior beliefs about the average change in the portfolio value μ by a Gaussian distribution, with parameters μ_0 and σ_0^2.

Bayesian Inference for the Gaussian Distribution

So in summary, we have that $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$. We know σ^2 (for now) and need to estimate μ.

The prior on μ is $p(\mu) = N(\mu_0, \sigma_0^2)$

It is **vital** to understand that σ^2 and σ_0^2 here denote two very, very different quantities. σ^2 is the variance Y_t (i.e. the variance of the percentage returns). σ_0^2 is a parameter of the prior distribution for μ and represents the uncertainty in our prior beliefs about μ.

Similarly, μ and μ_0 are different quantities. μ_0 is the mean of the prior distribution for μ and hence represents our prior beliefs about μ.

If you do not understand this then none of the following will make sense!

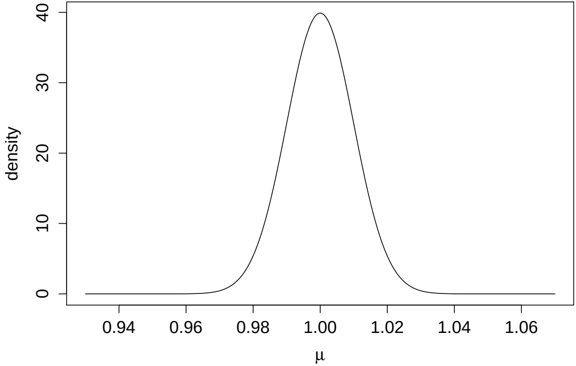
Example

Suppose a company has held a stock portfolio for 10 days. The observed percentage returns on these days are
0.997, 1.034, 1.012, 1.042, 1.017, 0.994, 1.040, 1.037, 1.022, 0.994

The company is interested in learning about the distribution of the percentage returns, for the purpose of computing the VaR and controlling risk. It assumes the percentage returns follow a Gaussian distribution $N(\mu, \sigma^2)$. Based on previous experience, the company knows that the true standard deviation of the returns is equal to $\sigma = 0.02$ (so $\sigma^2 = 0.0004$). It does not know the true mean μ, so it must estimate this. Based on its knowledge of other stock portfolios, it decides that its prior beliefs are best represented by a $p(\mu) \sim N(1, 0.01^2)$ prior – i.e. it believes the average percentage daily change is 1 (no change), and the standard deviation 0.01 measures the uncertainty in this prior belief

Example

The company's prior beliefs about μ are hence represented by a $N(1, 0.01^2)$ distribution.



Bayesian Inference for the Gaussian Distribution

As always, we use Bayes Theorem to derive the posterior distribution for μ after observing the data $Y = (Y_1, \ldots, Y_n)$: Note: we are conditioning on the variance σ^2 since this is assumed to be known.

$$p(\mu|Y, \sigma^2) = \frac{p(Y|\mu, \sigma^2)p(\mu)}{\int p(Y|\mu, \sigma^2)d\mu}$$

We know (since the prior is conjugate) that this will work out to be a Gaussian posterior $p(\mu|Y, \sigma^2) = N(\mu_1, \sigma_1^2)$ for some μ_1 and σ_1^2. The algebra is somewhat messy (although not difficult) so we will skip it - see supplementary lecture material on moodle if you are interested in the derivation

Bayesian Inference for the Gaussian Distribution

After doing the algebra, we find that $p(\mu|Y, \sigma^2) = N(\mu_1, \sigma_1^2)$ where:

$$\sigma_1^2 = \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}$$

$$\mu_1 = \frac{\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{t=1}^{n} Y_i}{\sigma^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)}$$

If we look closely at this, we can see that the posterior mean μ_1 is a weighted combination of the prior mean μ_0 and the empirical mean $\sum Y_t$, which intuitively makes sense. The posterior variance σ_1^2 is also lower than the prior variance σ_0^2 which again makes sense – we are more less uncertain about the value of μ after seeing the data than we were before.

Back to the Example

The company's prior for μ had $\mu_0 = 1$, $\sigma_0^2 = 0.01^2 = 0.0001$. The variance σ^2 was known to be 0.0004. The data was

$$Y = (0.997, 1.034, 1.012, 1.042, 1.017, 0.994, 1.040, 1.037, 1.022, 0.994)$$

which has empirical mean:

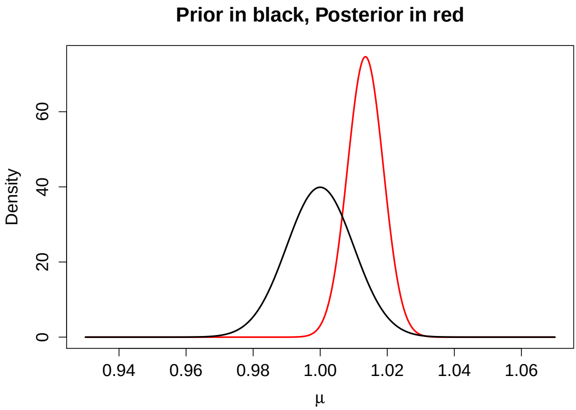
$$\frac{1}{10}\sum_{t=1}^{10} = 1.0189$$

Substituting all this into the previous formula gives $\mu_1 = 1.0135$ and $\sigma_1^2 = 0.0053^2$, so the posterior is:

$$p(\mu|Y, \sigma^2) = N(1.0135, 0.0053^2)$$

Example - Posterior

The company's posterior beliefs about μ can be plotted:



Note that this has shifted towards the empirical mean of the data (=1.0189) we we would expect

Non-informative Priors

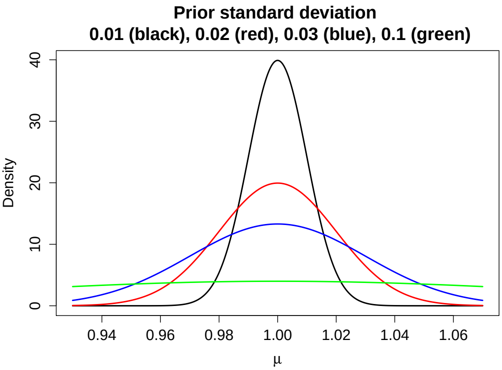
In practice we may not have strong prior beliefs about μ. Or we may not want our analysis to be influenced by our prior beliefs. In this case we can choose a prior that has an extremely large uncertainty and doesnt impose much prior information on the likelihood.

Recall that in Lecture 1 we saw something similar when we used a Beta(1,1) prior for θ in the Binomial distribution - this prior was a flat line and represented complete uncertainly about θ.

In the current case we can represent complete uncertainty about μ by making the prior variance parameter σ_0^2 very large. Remember - this is what controls how disperse the prior is, and hence how uncertain we are about μ before observing the data.

Noninformative Prior

Consider what happens when the prior on μ is $N(1, \sigma_0^2)$ and we increase the standard deviation σ



The prior is getting more and more disperse, corresponding to less and less information.

Non-informative Priors

So as σ^2 becomes larger and larger, the prior $p(\mu) = N(\mu_0, \sigma_0^2)$ becomes less and less informative. Ideally to represent complete ignorance we would take $\sigma_0^2 = \infty$ (or more formally, take the limit as σ_0^2 tends to infinity).

This leads to something called an improper prior – the prior is no longer a valid probability distribution. Does this matter? In this case no. The posterior works out to be well-defined.

Non-informative Priors

Recall we had:

$$\sigma_1^2 = \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right)^{-1}$$

$$\mu_1 = \frac{\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{t=1}^n Y_i}{\sigma^2} \right)}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right)}$$

Now let $\sigma_0^2 \to \infty$. The choice of μ_0 no longer matters, and this becomes:

$$\sigma_1^2 = \left(\frac{n}{\sigma^2} \right)^{-1} = \frac{\sigma^2}{n}$$

$$\mu_1 = \frac{\left(\frac{\sum_{t=1}^n Y_i}{\sigma^2} \right)}{\left(\frac{n}{\sigma^2} \right)} = \frac{\sum_{t=1}^n Y_i}{n}$$

Non-informative Priors

So if we have no strong prior beliefs about μ and use an uninformative prior, the posterior is simply:

$$p(\mu|Y, \sigma^2) = N\left(\frac{\sum_{t=1}^{n} Y_i}{n}, \frac{\sigma^2}{n}\right)$$

This makes intuitive sense – without any prior knowledge, our posterior mean is simply the empirical mean, and the posterior variance is the empirical variance.

So we can still do Bayesian inference even when we have no strong prior beliefs about the parameter (μ) here - just take prior that has a very high variance, like this one.

Back to the Example

Suppose the company in the previous example had no strong beliefs about μ and they used this non informative prior. Plugging in the numbers, their posterior would now be:

$$p(\mu|Y, \sigma^2) = N(1.0189, 0.0063^2)$$

(recall that 1.0189 was the empirical mean)

Previously, their posterior was $N(1.0135, 0.0053^2)$. This has been pulled away from the empirical mean, due to the prior belief that μ was around 1.

Note that the posterior variance when using the non informative prior is larger than when using the previous informative prior – starting out with more uncertainty means you end up with more uncertainty!

Computing the VaR

So, lets recap. We had historical returns $Y = (Y_1, \dots, Y_n)$. We assumed that their distribution was Gaussian N(μ, σ^2) with a known variance but unknown mean. We used Bayesian inference to find a posterior distribution for the mean.

Let Y_d denote the percentage change on an arbitrary day in the future which has the same distribution as the percentage changes in the historical sample, i.e $Y_d \sim N(\mu, \sigma^2)$. We don't know μ, but we have its posterior distribution $p(\mu | Y_1, \dots, Y_n)$

We now want to ask questions like "what is the probability of the portfolio losing more than 3% of its value on this day?". In the simple frequentist example we did earlier, we did not incorporate any uncertainty about μ when we done this - we simply used a point estimate, which led to a misleading risk calculation.

Computing the VaR

To answer "what is the probability of the portfolio losing more than 3% of its value on an arbitrary day?". we need $p(Y_d < 0.97)$

Similarly to answer "what is the probability of the portfolio losing more than 3% of its value on an arbitrary day?". we need $p(Y_d < 0.95)$

More generally, to answer "what is the probability of the portfolio losing more than Z% of its value on an arbitrary day?". we need $p(Y_d < Z)$

We are essentially trying to predict $p(Y_d < Z)$ based on the historical data $Y = (Y_1, \ldots, Y_n)$. We do this by using the historical data to find the posterior distribution for the unknown μ, and then predicting based on this.

In other words, we are interested in the distribution of Y_d based on incorporating information from the historical data. We write this distribution as $p(Y_d|Y)$

Computing the VaR

So, we have:

$$p(Y_d|\mu) = N(\mu, \sigma^2) \text{ (where } \sigma^2 \text{ is known}$$

$$p(\mu|Y) = N(\mu_1, \sigma_1^2) \text{ (posterior based on historic data } Y)$$

By the theorem of total probability we have:

$$p(Y_d|Y) = \int p(Y_d|\mu)p(\mu|Y)d\mu$$

This is the fundamental equation of Bayesian prediction.

Computing the VaR

It can be shown (for proof, see supplementary material on moodle) that in our conjugate Gaussian case:

$$p(Y_d|Y) = N(\mu_1, \sigma_1^2 + \sigma^2)$$

So to find

$$p(Y_d < Z|Y)$$

we need:

$$\int_{-\infty}^{Z} p(Y_d|Y)dY_d$$

which is just a Gaussian integral can do this in R using the pnorm() function.

Example

To return to our previous example, recall that the company (using the informative prior) had a posterior distribution

$$p(\mu|Y) = N(1.0135, 0.0053^2)$$

so $\mu_1 = 1.0135$ and $\sigma_1^2 = 0.0053^2$. The variance σ^2 was known to be 0.0004.

We hence have

$$p(Y_d|Y) = N(1.0135, 0.0053^2 + 0.0004) = N(1.0135, 0.000428)$$

Example

So if the company wants to know the probability of the portfolio dropping in value by more than 3% in a given day, this is:

$$\int_{-\infty}^{0.97} \frac{1}{\sqrt{2 \times 0.000428\pi}} e^{\left(\frac{(Y_d - 1.0135)^2}{2 \times 0.000428}\right)} dY_d$$

which can be done in R as:

```
> pnorm(0.97,1.0135, sqrt(0.000428))
[1] 0.01774811
```

so there is a 1.8% probability of the portfolio dropping in value by more than 3% in one day.

(Remember that the pnorm functioning R takes the standard deviation as an argument rather than the variance, hence why we square root)

Example

Similarly the probability of the portfolio dropping value by more than 5% in one day is:

```
> pnorm(0.95,1.0135, sqrt(0.000428))
[1] 0.001072488
```

so there is a 0.1% probability of the portfolio dropping in value by more than 5% in one day.

A Brief Note on Bayesian Inference for the Gaussian Distribution With Unknown Variance

In the previous analysis we assume the percentage change in portfolio returns Y_t has a $N(\mu, \sigma^2)$ distribution where σ^2 was known.

In practice σ^2 will usually be unknown. We must hence estimate both μ and σ^2.

For a Bayesian analysis, we need a prior $p(\mu, \sigma^2)$ on both parameters. Ideally, to make it easier to specify the prior, we perhaps want to treat the parameters as being independent and put a separate prior on each: $p(\mu, \sigma^2) = p(\mu)p(\sigma^2)$.

Unfortunately, this leads to a prior distribution that is not conjugate! We do not yet have the tools to work with non-conjugate priors, so we will leave this for now and revisit it in Lecture 5.