

Lecture 4: Value-at-Risk

Alex Donovan

1 February 2019

Value-at-Risk

- **Value-at-Risk** is probably the most widely used risk measure in financial institutions.
- It is used by banks, pensions funds, etc.
- 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

- **Value-at-Risk** is probably the most widely used risk measure in financial institutions.
- It is a measure for assessing the risk (i.e. the amount of potential loss) associated with an investment or a portfolio of investments.
- It is used by banks, pensions funds, etc.
- 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

- Value-at-Risk is the *market risk* measure prescribed by Basel Accord II and III to ensure that a bank has adequate capital to be able to absorb losses.
- Basel III is a set of international banking regulations established by the Basel Committee on Banking Supervision in order to promote stability in the international financial system.
- It builds on Basel I and Basel II, and was developed in response to the deficiencies in financial regulation revealed by the financial crisis of 2007-08

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 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

Let the random variable L denote the loss of a portfolio over the period h .

Definition 1.1 (Value-at-Risk)

$$\text{VaR}_{\alpha} = \inf \{l \in \mathbb{R} : P(L > l) \leq 1 - \alpha\} = \inf \{l \in \mathbb{R} : F_L(l) \geq \alpha\}$$

Given some confidence level $\alpha \in (0, 1)$, the Value-at-Risk of a given portfolio is the threshold such that the probability of losing more than this threshold over a time horizon h is equal to $1 - \alpha$.

Value at Risk - Definition

- VaR is hence a measure of **risk exposure**
- For example, if $\alpha = 0.99$ and the VaR is \$1m, this means that there is a 1% probability of losing more than \$1m in a single day.
- Normally, we will be given the portfolio, and the value α , then asked to find the VaR – i.e. for risk management the company will choose some value of α (e.g. $\alpha = 0.99$) then try to find what sum of money it has a $100(1 - \alpha)\%$ chance of losing in a single day.
- Typical values for α are $\alpha = 0.95$ or $\alpha = 0.99$.
- In *market risk* management the time horizon h is usually 1 or 10 days. In *credit risk* and *operational risk* management h is usually 1 year.

Value at Risk - Estimation

- There are many ways of computing the VaR, which have varying degrees of sophistication. We will consider a very simple method.
- In this course, rather than working with monetary loss, we will instead usually focus on **percentage loss**, since this is typically more informative.
- 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 is a random variable drawn from some distribution $F_Y(\cdot)$. 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. Suppose the company has held the portfolio for $n + 1$ days. They hence have n observations $\mathbf{y} = (y_1, \dots, y_n)$.
- For example, 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 $\$102m$.

Percentage Loss - Empirical Density

- The first 3 values of Y_t are then:

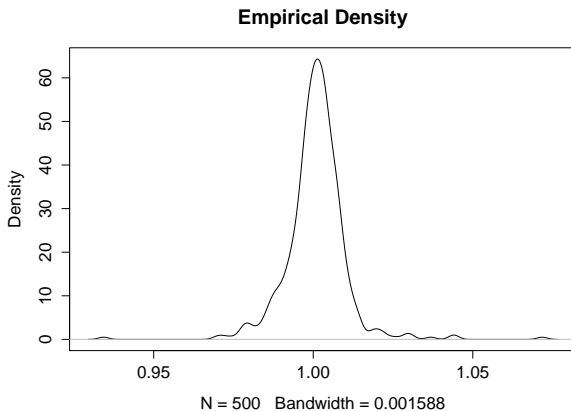
$$Y_1 = \frac{99}{100} = 0.99 \quad (\text{a } 1\% \text{ loss})$$

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

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

Values smaller than 1 denote a loss.

Percentage Loss - Empirical Density



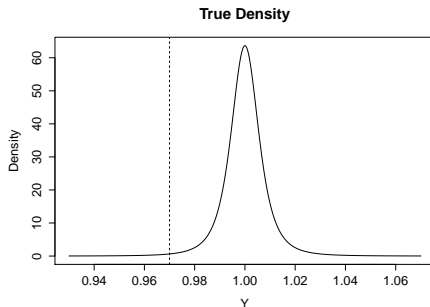
Value at Risk - Computation

- Suppose we know the true distribution $F_Y(\cdot)$ of Y_t . Then we can compute the VaR as follows:
- Find the $100(1 - \alpha)^{th}$ quantile of the **true** density of Y_t . Denote this by q .
- For example if $\alpha = 0.99$ then q is equal to the 1^{st} quantile¹, 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.

¹To be precise: 1^{st} 100-quantile or 1^{st} percentile

Percentage Loss - Computation

For example, if the following is the true distribution of Y_t and $\alpha = 0.99$.




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 possible 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 $\mathbf{y} = (y_1, \dots, y_n)$ 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 $\hat{\sigma}^2 = 1/n \sum_{t=1}^n (Y_t - \hat{\mu})^2$.
- For a given α , a simple estimate of q would be the value satisfying:

$$\int_{-\infty}^q p(y_t | \hat{\mu}, \hat{\sigma}^2) dy_t = (1 - \alpha)$$

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, i.e. $Y_i \sim N(1.01, 0.01)$.
- This means that on average the portfolio gains 1% in value each day, but the returns have a standard deviation of 1%.
- If $\alpha = 0.99$ then q can be easily computed using  software:
`qnorm(0.01, 1.01, 0.01) = 0.9867`
- So there is a 1% chance of the portfolio dropping by 1.33% or more in a given day (Note: $1 - 0.9867 = 0.0133$).

Value at Risk - Bayesian Approach

Several problems with this simple frequentist approach:

- 1 It does not take uncertainty about μ and σ^2 into account - we used point estimates (MLE). But these estimates will not be accurate - we could potentially underestimate risk because we are not taking this uncertainty into account.
- 2 It does not 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.

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 $p(\mu, \sigma^2 | y_1, \dots, y_n)$ given the historical data.
- 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 $p(\mu, \sigma^2 | \mathbf{y})$?
- 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 $p(y|\mu, \sigma^2)$.

Bayesian Inference for the Gaussian Distribution

- We now discuss how to perform Bayesian inference for parameters of the Gaussian distribution which describes the daily percentage changes in the portfolio value.
- Let's 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 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 .
- 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 $Y_1, \dots, Y_n \sim N(\mu, \sigma^2)$. 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 different quantities:
 - σ^2 is the variance of 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 .

Gaussian Distribution With Known Variance

Let's consider Bayesian estimation of the mean of a univariate Gaussian distribution whose variance is assumed to be known.

1 Likelihood

$$\begin{aligned} p(\mathbf{y}|\mu, \sigma^2) &= \prod_{i=1}^n p(y_i|\mu, \sigma^2) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right) \end{aligned} \quad (1)$$

Gaussian Distribution With Known Variance

$$\begin{aligned} p(\mathbf{y}|\mu, \sigma^2) &= \overbrace{\frac{1}{(2\pi)^{\frac{n}{2}}}}^{\text{constant}} \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2}\{ns^2 + n(\bar{y} - \mu)^2\}\right) \\ &\propto \underbrace{\frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{ns^2}{2\sigma^2}\right)}_{\text{constant}} \exp\left(-\frac{n}{2\sigma^2}(\bar{y} - \mu)^2\right) \end{aligned} \quad (2)$$

where $s^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$

When σ^2 is known, it can be treated as constant, and the above expression can be written as:

$$\begin{aligned} p(\mathbf{y}|\mu) &\propto \exp\left(-\frac{n}{2\sigma^2}(\bar{y} - \mu)^2\right) \\ &\propto \mathcal{N}(\bar{y}|\mu, \frac{\sigma^2}{n}) \end{aligned}$$

Gaussian Distribution With Known Variance

2 Prior

Since the likelihood has the form:

$$p(\mathbf{y}|\mu) \propto \exp\left(-\frac{n}{2\sigma^2}(\bar{y} - \mu)^2\right)$$

The *conjugate prior* will have the following form:

$$p(\mu) \propto \exp\left(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right)$$

That is, $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$:

$$p(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right)$$

Gaussian Distribution With Known Variance

3 Posterior Distribution

$$p(\mu|\mathbf{y}, \sigma^2) \propto p(\mu) \times p(\mathbf{y}|\mu, \sigma^2)$$

$$\propto \exp\left(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right) \times \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right)$$

$$= \exp\left(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right)$$

Gaussian Distribution With Known Variance

$$= \exp \left(-\frac{1}{2\sigma_0^2} (\mu^2 - 2\mu\mu_0 + \mu_0^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i^2 - 2y_i\mu + \mu^2) \right)$$

$$= \exp \left(-\frac{\mu^2}{2\sigma_0^2} + \frac{\mu\mu_0}{\sigma_0^2} - \frac{\mu_0^2}{2\sigma_0^2} - \frac{\sum_{i=1}^n y_i^2}{2\sigma^2} + \frac{\mu \sum_{i=1}^n y_i}{\sigma^2} - \frac{n\mu^2}{2\sigma^2} \right)$$

$$= \exp \left(-\frac{\mu^2}{2} \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right) + \mu \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^n y_i}{\sigma^2} \right) - \left(\frac{\mu_0^2}{2\sigma_0^2} + \frac{\sum_{i=1}^n y_i^2}{2\sigma^2} \right) \right)$$

Gaussian Distribution With Known Variance

Let's define new parameters σ_n^2 and μ_n :

$$\sigma_n^2 = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}} \quad (3)$$

$$\mu_n = \sigma_n^2 \left(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{y}}{\sigma^2} \right) \quad (4)$$

Gaussian Distribution With Known Variance

Then:

$$\begin{aligned} p(\mu|\mathbf{y}, \sigma^2) &\propto \exp\left(-\frac{1}{2\sigma_n^2}(\mu^2 - 2\mu\mu_n + \mu_n^2)\right) \\ &= \exp\left(-\frac{1}{2\sigma_n^2}(\mu - \mu_n)^2\right) \end{aligned} \quad (5)$$

Example

- Suppose a company has held a stock portfolio for 10 days.
- The observed returns on these days are:

$$\mathbf{y} = (0.997, 1.034, 1.012, 1.042, 1.017, 0.994, 1.040, 1.037, 1.022, 0.994)$$

- The company is interested in learning about the distribution of the returns, for the purpose of computing the VaR and controlling risk.

Example

- It is assumed that the percentage returns follow a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$.
- Based on previous experience, the company knows that the true standard deviation of the returns is equal to $\sigma = 0.02$ (so $\sigma^2 = 0.0004$).
- The company does not know the true mean μ , so it has to be estimated.
- Based on its knowledge of other stock portfolios, it decides that its prior beliefs are best represented by a prior $p(\mu) \sim \mathcal{N}(1, 0.01^2)$ – 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 (continued)

- Substituting information provided in the example into the equations (3) and (4) gives $\mu_n = \mu_{10} = 1.0135$ and $\sigma_n^2 = \sigma_{10}^2 = 0.0053^2$, so the posterior is:

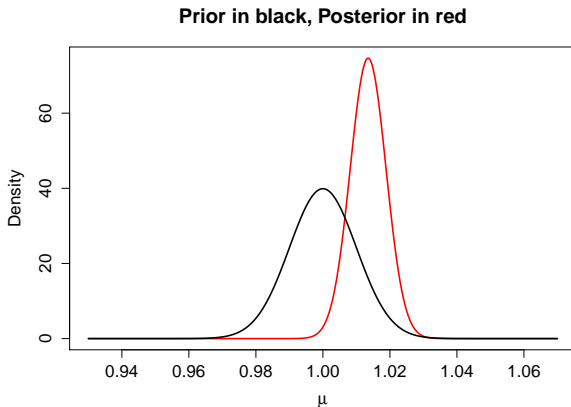
$$p(\mu|\mathbf{y}, \sigma^2) = \mathcal{N}(1.0135, 0.0053^2)$$

- The empirical mean of company's gross returns:

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

Example (continued)

The company's posterior beliefs about μ :



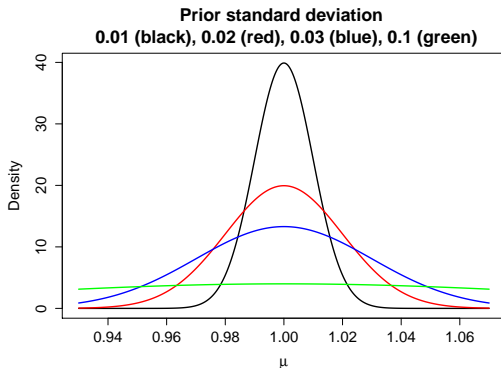
Note that the **posterior mean** has shifted towards the empirical mean of the data ($\bar{y} = 1.0189$).

Non-informative Priors

- In practice we may not have strong prior beliefs about μ .
- It could also be that we do not want our analysis to be influenced by our prior beliefs (e.g. $Beta(1,1)$).
- In this case we can choose a prior that has an extremely large uncertainty and doesn't impose much prior information on the likelihood.
- In the current case we can represent complete uncertainty about μ by making the prior variance parameter σ_0^2 very large.
- This parameter reflects how uncertain we are about μ before observing the data.

Non-informative Prior

Consider what happens when the prior on μ is $N(1, \sigma_0^2)$, and the variance σ_0^2 increases.



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

Non-informative Priors

- 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:

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

As we let $\sigma_0^2 \rightarrow \infty$, the choice of μ_0 no longer matters, and the above expressions can be written as follows:

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

Non-informative Priors

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

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

- 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 - just take prior that has a very high variance.

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:

$$p(\mu|\mathbf{y}, \sigma^2) = N(1.0189, 0.0063^2)$$

(recall that 1.0189 was the empirical mean)

- Previously, the 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 informative prior.

Computing the VaR

- Let \tilde{Y} denote the simple gross return on an arbitrary day, d , in the future, which has the same distribution as the gross returns in the historical sample, i.e $\tilde{Y} \sim N(\mu, \sigma^2)$.
- 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 not incorporate any uncertainty about μ when we computed VaR - we simply used a point estimate.

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 to compute $p(\tilde{Y} < 0.97)$.
- Similarly, to answer "what is the probability of the portfolio losing more than 5% of its value on an arbitrary day?", we need $p(\tilde{Y} < 0.95)$.

Posterior predictive distribution

- Let Y be a random variable with a known distribution $p(y|\theta)$, where θ is unknown.
- Before the data \mathbf{y} is observed, the marginal distribution of the unknown but observable y is as follows:

$$p(y) = \int p(y, \theta) d\theta = \int p(\theta) p(y|\theta) d\theta$$

- Once the data \mathbf{y} have been observed, we can predict an unknown \tilde{y} from the same process:

$$\begin{aligned} p(\tilde{y}|\mathbf{y}) &= \int p(\tilde{y}, \theta|\mathbf{y}) d\theta \\ &= \int p(\tilde{y}|\theta, \mathbf{y}) p(\theta|\mathbf{y}) d\theta \\ &= \int p(\tilde{y}|\theta) p(\theta|\mathbf{y}) d\theta \end{aligned}$$

Computing the VaR

- We are essentially trying to predict $p(\tilde{Y} < z)$ based on the historical data $\mathbf{y} = (y_1, \dots, y_n)$.
- We do this by using the historical data to find the posterior distribution for the unknown μ , and then predict based on this.
- In other words, we are interested in the distribution of \tilde{Y} based on incorporating information from the historical data. We write this distribution as $p(\tilde{y}|\mathbf{y})$.

Computing the VaR

- So, we have:

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

$$p(\mu|\mathbf{y}, \sigma^2) = N(\mu_n, \sigma_n^2) \text{ (posterior based on historic data } \mathbf{y})$$

- By the theorem of total probability we have:

$$p(\tilde{y}|\mathbf{y}) = \int p(\tilde{y}|\mu)p(\mu|\mathbf{y})d\mu$$

- This is the fundamental equation of Bayesian prediction.

Computing the VaR

- It can be shown that in our conjugate Gaussian case:

$$p(\tilde{y}|\mathbf{y}) = N(\mu_n, \sigma_n^2 + \sigma^2)$$

- In order to find:

$$p(\tilde{Y} < z|\mathbf{y})$$

we need:

$$\int_{-\infty}^z p(\tilde{y}|\mathbf{y}) d\tilde{y}$$

- which is just a Gaussian integral which can be easily computed.

Example

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

$$p(\mu|\mathbf{y}) = N(1.0135, 0.0053^2)$$

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

- Hence have:

$$\tilde{y}|\mathbf{y} \sim N(1.0135, 0.0053^2 + 0.0004) = N(1.0135, 0.000428)$$

Example

- If the company wants to know the probability of the portfolio dropping in value by more than 3% on a given day, we compute the following integral:

$$\int_{-\infty}^{0.97} \frac{1}{\sqrt{2 \times 0.000428 \pi}} e^{\left(\frac{(\tilde{y}-1.0135)^2}{2 \times 0.000428}\right)} d\tilde{y} = 0.01774811$$

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

Note on Bayesian Inference With Unknown Variance

- In the previous analysis we assumed that the gross returns Y_t had 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 later in the course.

Next week:

- Estimating the probability of extreme events occurring.
- We will look at techniques from the area of extreme value statistics.