The Bootstrap Method

version: 2018-02-09 · 18:14:04

Motivation: Quantile Regression

- **Normal Regression Model:** $y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \qquad \varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2).$
  - **MLE:** $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$
  - **Confidence Intervals:** $\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2\mathbf{V}), \quad \mathbf{V} = (\mathbf{X}'\mathbf{X})^{-1}.$

    $\implies$ 95% CI for β_j is $\hat{\beta}_j \pm 1.96 \cdot \hat{\sigma} V_{jj}^{1/2}$, where $\hat{\sigma}$ is the MLE of σ.

    (This is the Observed Fisher Information method, which is indistinguishable from the exact CI based on the $t_{(n-p)}$ distribution for $n - p > 30$.)

- **Relaxed Assumptions:** $y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \qquad \varepsilon_i \overset{\text{iid}}{\sim} f(\varepsilon),$

  $$E[\varepsilon_i] = 0, \ \text{var}(\varepsilon_i) = 1.$$

  - **Estimator:** Under Relaxed Assumptions, $\hat{\boldsymbol{\beta}}$ is the Best Linear Unbiased Estimator (BLUE), in the sense that for any $\tilde{\boldsymbol{\beta}} = \mathbf{A}\mathbf{y}$ with $E[\tilde{\boldsymbol{\beta}}] = \boldsymbol{\beta}$,

    $$\text{var}(\mathbf{a}'\hat{\boldsymbol{\beta}}) \le \text{var}(\mathbf{a}'\tilde{\boldsymbol{\beta}}), \qquad \mathbf{a} \in \mathbb{R}^p.$$

  - **Confidence Intervals:** By linearity still have $\text{var}(\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{V}$. Turns out that normality-based CI is asymptotically valid.

Motivation: Quantile Regression

- **Mean Regression:** $E[y \mid \mathbf{x}] = \mathbf{x}'\boldsymbol{\beta}$.
- **Quantile Regression:** Define the τ-level quantile function

  $$q_\tau(y \mid \mathbf{x}) = F^{-1}_{y \mid \mathbf{x}}(\tau \mid \mathbf{x}) \qquad \iff \qquad \Pr\{y \le q_\tau(y \mid \mathbf{x}) \mid \mathbf{x}\} = \tau.$$

  The QR model is

  $$q_\tau(y \mid \mathbf{x}) = \mathbf{x}'\boldsymbol{\beta},$$

  for any $\mathbf{x} \in \mathbb{R}^p$ and specific $\tau \in (0, 1)$ (or multiple τ each with their own $\boldsymbol{\beta}_\tau$).

Quantile Regression

Examples

1. **Additive Model:** $y = \mathbf{x}'\boldsymbol{\beta} + \varepsilon$, where ε is an arbitrary error independent of $\mathbf{x}$.

   $$\implies q_\tau(y \mid \mathbf{x}) = \mathbf{x}'\boldsymbol{\beta} + q_\tau(\varepsilon).$$

2. **Location-Scale Model:** $y = \mathbf{x}'\boldsymbol{\gamma} + \mathbf{x}'\boldsymbol{\eta} \cdot \varepsilon, \quad \varepsilon \amalg \mathbf{x}$.

   $$\implies q_\tau(y \mid \mathbf{x}) = \mathbf{x}'[\boldsymbol{\gamma} + \boldsymbol{\eta} \cdot q_\tau(\varepsilon)].$$

   (Having $\mathbf{x}$ in both mean and standard deviation is not a real restriction, i.e., set $\mathbf{x} = (\mathbf{z}, \mathbf{w})$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_z, \mathbf{0})$, $\boldsymbol{\eta} = (\mathbf{0}, \boldsymbol{\eta}_w)$.)

3. **Fixed-Quantile Error:** $y = \mathbf{x}'\boldsymbol{\beta} + \varepsilon$, where ε is not independent of $\mathbf{x}$, but

   $q_\tau(\varepsilon \mid \mathbf{x}) = \text{CONST}$.

4. **Fully specified QR model:** $q_\tau(y \mid \mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}_\tau$ for all $0 < \tau < 1$.

   Actually quite restrictive since quantiles need to be ordered:

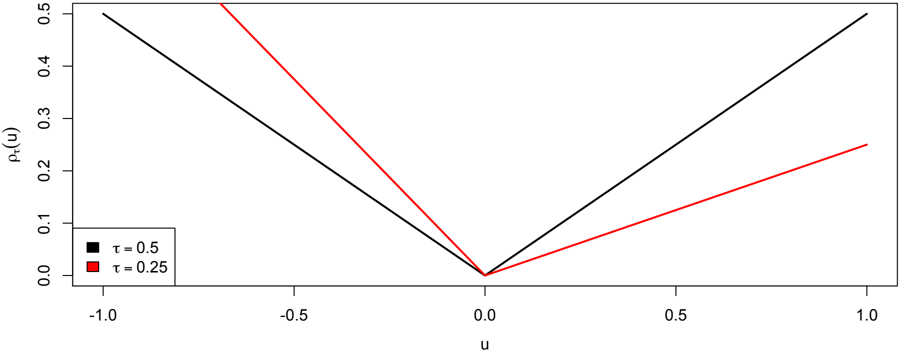
   $$\alpha_1 < \alpha_2 \qquad \implies \qquad \mathbf{x}'\boldsymbol{\beta}_{\alpha_1} < \mathbf{x}'\boldsymbol{\beta}_{\alpha_2} \quad \forall\, \mathbf{x} \in \mathbb{R}^p.$$

Parameter Estimation

- **Quantile Regression Model:** $q_\tau(y \mid \mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$ for given $\tau \in (0,1)$.
- **Moment Condition:** If true parameter value if $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, then

$$\boldsymbol{\beta}_0 = \arg\min_{\boldsymbol{\beta}} E[\rho_\tau(y - \mathbf{x}'\boldsymbol{\beta})], \qquad \rho_\tau(u) = u \cdot (\tau - \mathbb{1}\{u < 0\}).$$

- **Sample Analog:** $\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}_i'\boldsymbol{\beta})$.



Parameter Estimation

- **Quantile Regression Model:** $q_\tau(\alpha \mid \mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$ for given $\tau \in (0,1)$.
- **Point Estimate:**

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}_i'\boldsymbol{\beta}), \quad \rho_\tau(u) = u \cdot (\tau - \mathbb{1}\{u < 0\}).$$

  - **Equivalent Formulation:**

$$\min_{\boldsymbol{\beta}^+, \boldsymbol{\beta}^-, \mathbf{u}^+, \mathbf{u}^-} \sum_{i=1}^n \tau u_i^+ + (1-\tau) u_i^- \qquad \text{subject to} \qquad \mathbf{X}(\boldsymbol{\beta}^+ - \boldsymbol{\beta}^-) + \mathbf{u}^+ - \mathbf{u}^- = \mathbf{y},$$

  where $\beta_j^+ = \max(\beta_j, 0)$, $\beta_j^- = -\min(\beta_j, 0)$ and similarly for u_i^+ and u_i^-.

  - This is a **linear program** in $\mathbf{w} = (\boldsymbol{\beta}^+, \boldsymbol{\beta}^-, \mathbf{u}^+, \mathbf{u}^-)$,

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathbf{c}'\mathbf{w} \qquad \text{subject to} \qquad \mathbf{A}\mathbf{w} \le \mathbf{b}, \mathbf{w} \ge \mathbf{0},$$

  for which efficient algorithms are available.

Quantile Regression

- **Model:** $q_\tau(y \mid \mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$ for given $\tau \in (0,1)$.
- **Point Estimate:** $\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}_i'\boldsymbol{\beta})$ via linear programming.
- **Confidence Intervals:** ???, since we don't have a likelihood to calculate Observed Fisher Information!
  - Add modeling assumptions $\implies \hat{\boldsymbol{\beta}} \to \mathcal{N}(\boldsymbol{\beta}_0, \boldsymbol{\Sigma})$, but $\boldsymbol{\Sigma}$ is difficult to estimate (nonparametric smoothing estimator with high variance).
  - Can do something much simpler... (but computationally more intensive)

The Bootstrap Method

- **Setup:**
  - **Data and Model:** $\mathbf{y} = (y_1, \ldots, y_n) \sim F(\mathbf{y})$.

    I.e., completely general data-generating process (DGP) on the random vector $\mathbf{y}$. Could be a parametric model $\mathbf{y} \sim f(\mathbf{y} \mid \boldsymbol{\theta})$, a nonparametric model $y_i \overset{\text{iid}}{\sim} F(y)$, or a semi-parametric model like quantile regression...
  - **Quantity of Interest:** $\tau_0 = \mathcal{G}(F)$.

    I.e., τ_0 must be some functional of the DGP. Could be $\tau_0 = \tau(\boldsymbol{\theta}_0)$, or the median of F...
  - **Estimator:** $\hat{\tau} = g(\mathbf{y})$.
- **Objective:** Calculate a confidence interval for τ_0.

The Bootstrap Method

- **Setup:**
  - **Data and Model:** $\mathbf{y} = (y_1, \dots, y_n) \sim F(\mathbf{y})$.
  - **Quantity of Interest:** $\tau_0 = \mathcal{G}(F)$.
  - **Estimator:** $\hat{\tau} = g(\mathbf{y})$.
- **Objective:** Calculate a confidence interval for τ_0.
- **Problem:** Can't use likelihood theory because:

1. Sample size n is too small for asymptotics to kick in.
2. Don't have a parametric likelihood $f(\mathbf{y} \mid \boldsymbol{\theta})$.
3. Have likelihood but estimator is not MLE (e.g., lasso for variable selection).
4. Have likelihood + MLE, but suspect some degree of model misspecification.

The Bootstrap Method

- **Data and Model:** $\mathbf{y} = (y_1, \dots, y_n) \sim F(\mathbf{y})$.
- **Quantity of Interest:** $\tau_0 = \mathcal{G}(F)$.
- **Estimator:** $\hat{\tau} = g(\mathbf{y})$.
- **Objective:** Calculate a confidence interval for τ_0.
- **Idealized Scenario:** Suppose an oracle gives you the distribution of the *pivotal quantity* $T = \tau_0 - \hat{\tau}$. Then

$$\Pr(L < T < U) = \Pr(L < \tau_0 - \hat{\tau} < U) = \Pr(\hat{\tau} + L < \theta_0 < \hat{\tau} + U).$$

$\implies$ If L/U are the 2.5/97.5% quantiles of T, then a 95% CI for τ_0 is

$$\tau_0 \in (\hat{\tau} + L, \hat{\tau} + U).$$

The Bootstrap Method

- **Data and Model:** $\mathbf{y} = (y_1, \ldots, y_n) \sim F(\mathbf{y})$.
- **Quantity of Interest:** $\tau_0 = \mathcal{G}(F)$.
- **Estimator:** $\hat{\tau} = g(\mathbf{y})$.
- **Oracle:** Suppose distribution of $T = \tau_0 - \hat{\tau}$ is given.

  If L/U are the 2.5/97.5% quantiles of T, then CI for τ_0 is $(\hat{\tau} + L, \hat{\tau} + U)$.
- **Bootstrap:** Estimate L and U as follows:
  1. **Simulate** M datasets $\tilde{\mathbf{y}}^{(m)} \stackrel{\text{iid}}{\sim} \hat{F}(\mathbf{y})$, each of size n, where $\hat{F}(\mathbf{y})$ is an estimate of $F(\mathbf{y})$. The two most common ways to do this are:
     - **i. Parametric Bootstrap:** If $\mathbf{y} \sim f(\mathbf{y} \mid \boldsymbol{\theta})$, then $\tilde{\mathbf{y}}^{(m)} \stackrel{\text{iid}}{\sim} f(\mathbf{y} \mid \hat{\boldsymbol{\theta}})$.
     - **ii. Nonparametric Bootstrap:** If $y_i \stackrel{\text{iid}}{\sim} F(y)$, then $y_i^{(m)} \stackrel{\text{iid}}{\sim} \hat{F}(y)$, where $\hat{F}(y)$ is the empirical CDF of $\mathbf{y}$. In other words, $\tilde{\mathbf{y}}^{(m)}$ is sampled n times with replacement from $\mathbf{y}$.

The Bootstrap Method

- **Data and Model:**
  $\mathbf{y} = (y_1, \dots, y_n) \sim F(\mathbf{y})$.
- **Quantity of Interest:** $\tau_0 = \mathcal{G}(F)$.
- **Estimator:** $\hat{\tau} = g(\mathbf{y})$.
- **Oracle:** Suppose distribution of $T = \tau_0 - \hat{\tau}$ is given.

  If L/U are the 2.5/97.5% quantiles of T, then CI for τ_0 is $(\hat{\tau} + L, \hat{\tau} + U)$.
- **Bootstrap:** Estimate L and U as follows:
  1. **Simulate** M datasets $\tilde{\mathbf{y}}^{(m)} \overset{\text{iid}}{\sim} \hat{F}(\mathbf{y})$, each of size n, where $\hat{F}(\mathbf{y})$ is an estimate of $F(\mathbf{y})$.
  2. For each dataset, calculate $\tilde{\tau}^{(m)} = g(\tilde{\mathbf{y}}^{(m)})$ and $\tilde{T}^{(m)} = \hat{\tau} - \tilde{\tau}^{(m)}$.
  3. Let $\tilde{L}/\tilde{U}$ be the 2.5/97% *sample* quantiles of $\tilde{T}^{(1)}, \dots, \tilde{T}^{(M)}$.

  $\implies$ The Bootstrap CI for τ_0 is given by $(\hat{\tau} + \tilde{L}, \hat{\tau} + \tilde{U})$.

The Bootstrap Method

	Real World	Bootstrap World
Sampling Distribution	$\mathbf{y} \sim F(\mathbf{y})$	$\tilde{\mathbf{y}} \sim \hat{F}(\mathbf{y})$
Quantity of Interest	$\tau_0 = \mathcal{G}(F)$	$\hat{\tau} = g(\mathbf{y})$
Estimator	$\hat{\tau} = g(\mathbf{y})$	$\tilde{\tau} = g(\tilde{\mathbf{y}})$
Pivotal Quantity	$T = \tau_0 - \hat{\tau}$	$\tilde{T} = \hat{\tau} - \tilde{\tau}$
Quantiles:	$P(L < T < U) = 95\%$	$P(\tilde{L} < \tilde{T} < \tilde{U}) = 95\%$
95% Confidence Interval	Oracle: $(\hat{\tau} + L, \hat{\tau} + U)$ Bootstrap: $(\hat{\tau} + \tilde{L}, \hat{\tau} + \tilde{U})$	

Parallel between the Real world and the Bootstrap world.

Example: Range of Uniform

- **Objective:** Given $\mathbf{U} = (U_1, \dots, U_n)$, $U_i \overset{\text{iid}}{\sim} \text{Unif}(0, \theta)$, we wish to estimate θ.
- **Simulation Study:** Generate $N = 1000$ datasets with $\theta_0 = 1$ and perform calculations for each of the following settings:

  1. **Sample Size:** (i) $n = 100$ and (ii) $n = 10000$
  2. **Estimators:** (i) MLE $\hat{\theta}_1 = \max(\mathbf{U})$ and (ii) Unbiased $\hat{\theta}_2 = 2\bar{\mathbf{U}}$ (since $E[\bar{\mathbf{U}}] = \theta/2$).
  3. **Bootstrap Sampling:** (i) Nonparametric ($\tilde{\mathbf{U}}$ sampled with replacement) and (ii) Parametric ($\tilde{U}_i^{(m)} \overset{\text{iid}}{\sim} \text{Unif}(0, \hat{\theta})$). Always use $M = 1000$ bootstrap samples.
  4. **Confidence Intervals:** (i) Basic Bootstrap: $(\hat{\theta} + \tilde{L}, \hat{\theta} + \tilde{U})$
     (ii) Percentile Bootstrap: 2.5/97.5% quantiles of $\tilde{\theta}^{(1)}, \dots \tilde{\theta}^{(M)}$. (seems simpler but...)
  5. **Model Misspecification:** True sampling distribution is $U_i \overset{\text{iid}}{\sim} \theta \times \text{Beta}(\alpha, \alpha)$, where (i) $\alpha = 1$ ($\text{Beta}(1,1) = \text{Unif}(0,1)$) and (ii) $\alpha = 2$.
     (θ is range of distribution, so still meaningfull quantity to estimate. For $\alpha \neq 1$, $\hat{\theta}_1$ no longer MLE, but $\hat{\theta}_2$ still unbiased.)

- **Comparison Metrics:** (i) True coverage of CI and (ii) Average width of CI.

Example: Range of Uniform

- **Objective:** Given $\mathbf{U} = (U_1, \ldots, U_n)$, $U_i \overset{\text{iid}}{\sim} \text{Unif}(0, \theta)$, wish to estimate θ.
- **Simulation Study:** For $N = 1000$ datasets with $\theta_0 = 1$:
  1. **Sample Size:** (i) $n = 100$ and (ii) $n = 10000$
  2. **Estimators:** (i) $\hat{\theta}_1 = \max(\mathbf{U})$ and (ii) $\hat{\theta}_2 = 2\bar{\mathbf{U}}$.
  3. **Bootstrap Sampling:** For $M = 1000$ bootstrap samples, sampling is
     - **i.** Nonparametric: $\tilde{\mathbf{U}}$ sampled with replacement.

       Variance Reduction: Use same $\tilde{\mathbf{U}}^{(m)}$ to calculate both $\hat{\theta}_1^{(m)}$ and $\hat{\theta}_2^{(m)}$.
       $\implies$ Monte Carlo *difference* between comparison metrics has same expectation, but lower variance
     - **ii.** Parametric: $\tilde{U}_i^{(m)} \overset{\text{iid}}{\sim} \text{Unif}(0, \hat{\theta})$.
  4. **Confidence Intervals:** (i) Basic Bootstrap and (ii) Percentile Bootstrap.
  5. **Model Misspecification:** $U_i \overset{\text{iid}}{\sim} \theta \times \text{Beta}(\alpha, \alpha)$, where (i) $\alpha = 1$ and (ii) $\alpha = 2$.
- **Comparison Metrics:** (i) True coverage of CI and (ii) Average width of CI.

Example: Range of Uniform

- **Objective:** Given $\mathbf{U} = (U_1, \dots, U_n)$, $U_i \overset{\text{iid}}{\sim} \text{Unif}(0, \theta)$, wish to estimate θ.
- **Simulation Study:** For $N = 1000$ datasets with $\theta_0 = 1$:

1. **Sample Size:** (i) $n = 100$ and (ii) $n = 10000$
2. **Estimators:** (i) $\hat{\theta}_1 = \max(\mathbf{U})$ and (ii) $\hat{\theta}_2 = 2\bar{\mathbf{U}}$.
3. **Bootstrap Sampling:** For $M = 1000$ bootstrap samples, sampling is
   i. Nonparametric: $\tilde{\mathbf{U}}$ sampled with replacement.

      Variance Reduction: Use same $\tilde{\mathbf{U}}^{(m)}$ to calculate both $\hat{\theta}_1^{(m)}$ and $\hat{\theta}_2^{(m)}$.

   ii. Parametric: $\tilde{U}_i^{(m)} \overset{\text{iid}}{\sim} \text{Unif}(0, \hat{\theta})$.

      Variance Reduction: Use same $\tilde{R}_i^{(m)} \overset{\text{iid}}{\sim} \text{Unif}(0, 1)$, and let $\tilde{U}_i^{(m)} = \hat{\theta}_k \tilde{R}_i^{(m)}$, $k = 1, 2$.
4. **Confidence Intervals:** (i) Basic Bootstrap and (ii) Percentile Bootstrap.
5. **Model Misspecification:** $U_i \overset{\text{iid}}{\sim} \theta \times \text{Beta}(\alpha, \alpha)$, where (i) $\alpha = 1$ and (ii) $\alpha = 2$.

- **Comparison Metrics:** (i) True coverage of CI and (ii) Average width of CI.

Example: Range of Uniform

Actual Coverage

alpha = 1

	NP_max	NP_mean2	P_max	P_mean2
basic_n=100	0.86	0.94	0.94	0.95
basic_n=10K	0.89	0.95	0.95	0.94
pct_n=100	0.00	0.95	0.00	0.95
pct_n=10K	0.00	0.95	0.00	0.95

alpha = 2

	NP_max	NP_mean2	P_max	P_mean2
basic_n=100	0.60	0.93	0.29	0.98
basic_n=10K	0.57	0.95	0.00	0.98
pct_n=100	0.00	0.94	0.00	0.98
pct_n=10K	0.00	0.95	0.00	0.99

Interval Width

alpha = 1

	NP_max	NP_mean2	P_max	P_mean2
basic_n=100	0.03	0.22	0.04	0.22
basic_n=10K	0.00	0.02	0.00	0.02
pct_n=100	0.03	0.22	0.04	0.22
pct_n=10K	0.00	0.02	0.00	0.02

alpha = 2

	NP_max	NP_mean2	P_max	P_mean2
basic_n=100	0.07	0.17	0.03	0.22
basic_n=10K	0.01	0.02	0.00	0.02
pct_n=100	0.07	0.17	0.03	0.22
pct_n=10K	0.01	0.02	0.00	0.02

Remarks:

1. Percentile CI based on $\hat{\theta}_1 = \max(\mathbf{U})$ has 0% coverage! This is because $\theta_0 > \hat{\theta}_1 > \tilde{\theta}_1^{(m)}$, so quantiles of $\tilde{\theta}_1^{(m)}$ can never cover θ_0.

Example: Range of Uniform

Actual Coverage

alpha = 1

	NP_max	NP_mean2	P_max	P_mean2
basic_n=100	0.86	0.94	0.94	0.95
basic_n=10K	0.89	0.95	0.95	0.94
pct_n=100	0.00	0.95	0.00	0.95
pct_n=10K	0.00	0.95	0.00	0.95

alpha = 2

	NP_max	NP_mean2	P_max	P_mean2
basic_n=100	0.60	0.93	0.29	0.98
basic_n=10K	0.57	0.95	0.00	0.98
pct_n=100	0.00	0.94	0.00	0.98
pct_n=10K	0.00	0.95	0.00	0.99

Interval Width

alpha = 1

	NP_max	NP_mean2	P_max	P_mean2
basic_n=100	0.03	0.22	0.04	0.22
basic_n=10K	0.00	0.02	0.00	0.02
pct_n=100	0.03	0.22	0.04	0.22
pct_n=10K	0.00	0.02	0.00	0.02

alpha = 2

	NP_max	NP_mean2	P_max	P_mean2
basic_n=100	0.07	0.17	0.03	0.22
basic_n=10K	0.01	0.02	0.00	0.02
pct_n=100	0.07	0.17	0.03	0.22
pct_n=10K	0.01	0.02	0.00	0.02

Remarks:

2. NP bootstrap with Basic CI does not approach 95% coverage as sample size $n \to \infty$! This is because bootstrap only works if $\tilde{\theta}$ and $\hat{\theta}$ have the same distribution as $n \to \infty$. However, $\hat{\theta}_1 \sim \theta_0 \times \text{Beta}(1, n)$ is a continuous distribution, but

$$\Pr(\tilde{\theta}_1 = \hat{\theta}_1) = 1 - \Pr(\tilde{\theta}_1 \neq \hat{\theta}_1) = 1 - (1 - \tfrac{1}{n})^n \to 1 - e^{-1} \approx 0.63.$$

Therefore, $\tilde{\theta}_1$ has a non-vanishing point mass at $\hat{\theta}_1$, so doesn't get close to continuous distribution of $\hat{\theta}_1$.

Example: Range of Uniform

Actual Coverage

alpha = 1

	NP_max	NP_mean2	P_max	P_mean2
basic_n=100	0.86	0.94	0.94	0.95
basic_n=10K	0.89	0.95	0.95	0.94
pct_n=100	0.00	0.95	0.00	0.95
pct_n=10K	0.00	0.95	0.00	0.95

alpha = 2

	NP_max	NP_mean2	P_max	P_mean2
basic_n=100	0.60	0.93	0.29	0.98
basic_n=10K	0.57	0.95	0.00	0.98
pct_n=100	0.00	0.94	0.00	0.98
pct_n=10K	0.00	0.95	0.00	0.99

Interval Width

alpha = 1

	NP_max	NP_mean2	P_max	P_mean2
basic_n=100	0.03	0.22	0.04	0.22
basic_n=10K	0.00	0.02	0.00	0.02
pct_n=100	0.03	0.22	0.04	0.22
pct_n=10K	0.00	0.02	0.00	0.02

alpha = 2

	NP_max	NP_mean2	P_max	P_mean2
basic_n=100	0.07	0.17	0.03	0.22
basic_n=10K	0.01	0.02	0.00	0.02
pct_n=100	0.07	0.17	0.03	0.22
pct_n=10K	0.01	0.02	0.00	0.02

Remarks:

3. NP-CI for $\hat{\theta}_2$ have the right coverage, even under wrong model $\alpha = 2$. On the other hand P-CI with $\hat{\theta}_2$ overcover under wrong model (98% instead of 95%).

Example: Range of Uniform

Actual Coverage

alpha = 1

	NP_max	NP_mean2	P_max	P_mean2
basic_n=100	0.86	0.94	0.94	0.95
basic_n=10K	0.89	0.95	0.95	0.94
pct_n=100	0.00	0.95	0.00	0.95
pct_n=10K	0.00	0.95	0.00	0.95

alpha = 2

	NP_max	NP_mean2	P_max	P_mean2
basic_n=100	0.60	0.93	0.29	0.98
basic_n=10K	0.57	0.95	0.00	0.98
pct_n=100	0.00	0.94	0.00	0.98
pct_n=10K	0.00	0.95	0.00	0.99

Interval Width

alpha = 1

	NP_max	NP_mean2	P_max	P_mean2
basic_n=100	0.03	0.22	0.04	0.22
basic_n=10K	0.00	0.02	0.00	0.02
pct_n=100	0.03	0.22	0.04	0.22
pct_n=10K	0.00	0.02	0.00	0.02

alpha = 2

	NP_max	NP_mean2	P_max	P_mean2
basic_n=100	0.07	0.17	0.03	0.22
basic_n=10K	0.01	0.02	0.00	0.02
pct_n=100	0.07	0.17	0.03	0.22
pct_n=10K	0.01	0.02	0.00	0.02

Remarks:

4. $\hat{\theta}_1$ does not converge to θ_0 under the wrong model $\alpha = 2$, so CI has poor coverage. On the other hand, interval width is narrower than with $\hat{\theta}_2$, because max has less variance than mean.

Example: GARCH Stochastic Volatility Model

- **SDE SV Model:** Let $(\Delta X_t, \Delta V_t)$ be the asset/volatility log-return/return on day t. The basic SDE-SV model is

$$\Delta X_t = (\alpha - \tfrac{1}{2} V_t)\Delta t + V_t^{1/2} \Delta B_{1t}$$
$$\Delta V_t = -\gamma(V_t - \mu)\Delta t + \sigma V_t^{1/2} \Delta B_{2t}$$

  - Pros: Excellent performance; easy to calibrate when V_t is observed (e.g., VIX for GSPC).
  - Cons: Extremely difficult to calibrate when V_t is latent, since $\ell(\boldsymbol{\theta} \mid \mathbf{X})$ is not available in closed-form, i.e,

$$\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{X}) \propto p(\mathbf{X} \mid \boldsymbol{\theta}) = \int p(\mathbf{X}, \mathbf{V} \mid \boldsymbol{\theta})\, \mathrm{d}\mathbf{V}$$

GARCH Stochastic Volatility Model

- **SDE SV Model:**

$$\Delta X_t = (\alpha - \tfrac{1}{2} V_t)\Delta t + V_t^{1/2} \Delta B_{1t}$$
$$\Delta V_t = -\gamma(V_t - \mu)\Delta t + \sigma V_t^{1/2} \Delta B_{2t}$$

- **GARCH SV Model:** Let $\varepsilon_t = \Delta X_t$. The GARCH(1,1) model is

$$\varepsilon_t = \sigma_t z_t, \qquad z_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$$
$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

  - Like SDEs, volatility σ_t is stochastic.
  - Pros: Inference with GARCH is far simpler than with SDE (closed-form likelihood).
  - Cons: Unlike SDEs, GARCH is a discrete-time model (difficult for option pricing and consistency across timescales)

GARCH Stochastic Volatility Model

- **Data:** Asset values $\mathbf{S} = (S_0, \ldots, S_N) \implies$ log-returns $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_N)$, with $\varepsilon_t = \log(S_t / S_{t-1})$.
- **GARCH(1,1) Model:**
$$\varepsilon_t = \sigma_t z_t, \qquad z_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$$
$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$
- **Objective:** On given day N, estimate the p-day forward τ-level Value-At-Risk, i.e., the conditional quantile

$$\mathsf{VaR}_\tau = q_\tau \left(\frac{S_{N+p} - S_N}{S_N} \,|\, \mathbf{S}, \boldsymbol{\theta} \right) \quad \iff \quad \Pr\left(\frac{S_{N+p} - S_N}{S_N} < \mathsf{VaR}_\tau \,|\, \mathbf{S}, \boldsymbol{\theta} \right) = \tau.$$

For example, we would say that the 10-day 5%-level VaR of AAPL is a 1.3% drop in value.

GARCH Model

- **Model:** $$\varepsilon_t = \sigma_t z_t, \qquad z_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$$
  $$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$
- **Parameter Estimation:**
  - **R** *Packages:* `rugarch`, `fGarch`. The former is more stable, the latter is faster. Both can fit numerous extensions to the basic GARCH(1,1) model above.
  - *Profile Likelihood:* For $\boldsymbol{\theta} = (\omega, \alpha, \beta)$

$$\begin{aligned} \ell(\boldsymbol{\theta} \mid \boldsymbol{\varepsilon}) &= -\frac{1}{2}\sum_{t=1}^N \frac{\varepsilon_t^2}{\sigma_t^2} + \log(\sigma_t^2), & \sigma_t^2 &= \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \\ &= -\frac{1}{2}\sum_{t=1}^N \frac{\varepsilon_t^2}{\omega \cdot \tilde{\sigma}_t^2} + \log(\omega \cdot \tilde{\sigma}_t^2), & \tilde{\sigma}_t^2 &= 1 + \eta\varepsilon_{t-1}^2 + \beta\tilde{\sigma}_{t-1}^2, \end{aligned}$$

where $\eta = \alpha/\omega \implies \hat{\omega}(\eta, \beta) = \sum_{t=1}^N (\varepsilon_t/\tilde{\sigma}_t)^2$.

(Note the technical issue of initializing $\tilde{\sigma}_1$ which we won't discuss here.)

Value-at-Risk

- **GARCH(1,1) Model:**
$$\varepsilon_t = \sigma_t z_t, \qquad z_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$$
$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

- **Value-at-Risk:**
$$\mathrm{VaR}_\tau = q_\tau\left(\frac{S_{N+p} - S_N}{S_N} \,\Big|\, \mathbf{S}, \boldsymbol{\theta}\right) \quad \iff \quad \Pr\left(\frac{S_{N+p} - S_N}{S_N} < \mathrm{VaR}_\tau \,\Big|\, \mathbf{S}, \boldsymbol{\theta}\right) = \tau.$$

- **1-Day VaR**: For given $\boldsymbol{\theta}$ and data $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_N)$

  1. Let $\sigma_1^2 = E[\sigma_1^2 \mid \boldsymbol{\theta}] = \omega/(1-\alpha-\beta)$

  2. Use GARCH equation to obtain $\sigma_{N+1}^2 = \omega + \alpha\varepsilon_N^2 + \beta\sigma_N^2$

  3. $(S_{N+1} - S_N)/S_N = \exp(\varepsilon_{N+1}) - 1 \implies \mathrm{VaR}_\tau = \exp\{\mathtt{qnorm}(\tau \mid 0, \sigma_{N+1})\} - 1$

  In other words, $\mathrm{VaR}_\tau = \mathrm{VaR}_\tau(\boldsymbol{\theta} \mid \boldsymbol{\varepsilon})$ is a function of $\boldsymbol{\theta}$ (and observed data $\boldsymbol{\varepsilon}$).

Value-at-Risk

- **GARCH(1,1) Model:**
$$\varepsilon_t = \sigma_t z_t, \qquad z_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$$
$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

- **1-Day VaR**: For given $\boldsymbol{\theta}$ and data $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_N)$

  1. Let $\sigma_1^2 = E[\sigma_1^2 \mid \boldsymbol{\theta}] = \omega/(1-\alpha-\beta)$

  2. Use GARCH equation to obtain $\sigma_{N+1}^2 = \omega + \alpha\varepsilon_N^2 + \beta\sigma_N^2$

  3. $(S_{N+1} - S_N)/S_N = \exp(\varepsilon_{N+1}) - 1 \implies \mathrm{VaR}_\tau = \exp\{\mathtt{qnorm}(\tau \mid 0, \sigma_{N+1})\} - 1$

  In other words, $\mathrm{VaR}_\tau = \mathrm{VaR}_\tau(\boldsymbol{\theta} \mid \boldsymbol{\varepsilon})$.

- **Inference:** If $\hat{\boldsymbol{\theta}}$ is the MLE of GARCH model, then
  - **MLE:** Use plug-in principle: $\widehat{\mathrm{VaR}}_\tau = \mathrm{VaR}_\tau(\hat{\boldsymbol{\theta}} \mid \boldsymbol{\varepsilon})$.
  - **Confidence Intervals?**

Delta-Method

- **Setup:**
  - Let $Y_1, Y_2, \dots$ be some stochastic process determined by a parameter $\boldsymbol{\theta} \in \mathbb{R}^p$ (In the simplest case, we have $Y_n \overset{\text{iid}}{\sim} f(y \mid \boldsymbol{\theta})$, but the theory works for stationary processes such as GARCH(1,1) as well).
  - For $\mathbf{Y}_{1:n} = (Y_1, \dots, Y_n)$, suppose the MLE and the inverse Fisher Information

$$\hat{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta} \mid \mathbf{Y}_{1:n}), \qquad \hat{\mathbf{V}}_n = \left[-\tfrac{\partial^2}{\partial \boldsymbol{\theta}^2} \ell(\boldsymbol{\theta} \mid \mathbf{Y}_{1:n})\right]^{-1}$$

    satisfy the usual asymptotic theory, i.e., $\hat{\mathbf{V}}_n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \to \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$ as $n \to \infty$, where $\boldsymbol{\theta}_0$ is the true parameter value.

- **Theorem:** Suppose that $\boldsymbol{\tau} : \mathbb{R}^p \to \mathbb{R}^q$ is a differentiable function with $q \leq p$, and we wish to estimate $\boldsymbol{\tau}_0 = \boldsymbol{\tau}(\boldsymbol{\theta}_0)$. Then as $n \to \infty$ we have

$$\hat{\boldsymbol{\Sigma}}_n^{1/2}(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0) \to \mathcal{N}(\mathbf{0}, \mathbf{I}), \qquad \hat{\boldsymbol{\tau}}_n = \boldsymbol{\tau}(\hat{\boldsymbol{\theta}}_n)$$

$$\hat{\boldsymbol{\Sigma}}_n = [\nabla \boldsymbol{\tau}(\hat{\boldsymbol{\theta}}_n)]' \hat{\mathbf{V}}_n [\nabla \boldsymbol{\tau}(\hat{\boldsymbol{\theta}}_n)].$$

Delta-Method

- **Setup:** Suppose that as $n \to \infty$ we have $\hat{\mathbf{V}}_n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \to \mathcal{N}(\mathbf{0}, \mathbf{I})$, where $(\hat{\boldsymbol{\theta}}_n, \hat{\mathbf{V}}_n)$ are the MLE and inverse Fisher Information calculated from a sequence of random variables $\mathbf{Y}_{1:n}$.

- **Theorem:** Suppose that $\boldsymbol{\tau} : \mathbb{R}^p \to \mathbb{R}^q$ is a differentiable function with $q \le p$, and we wish to estimate $\boldsymbol{\tau}_0 = \boldsymbol{\tau}(\boldsymbol{\theta}_0)$. Then as $n \to \infty$ we have

$$\hat{\boldsymbol{\Sigma}}_n^{1/2}(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0) \to \mathcal{N}(\mathbf{0}, \mathbf{I}), \qquad \begin{aligned} \hat{\boldsymbol{\tau}}_n &= \boldsymbol{\tau}(\hat{\boldsymbol{\theta}}_n) \\ \hat{\boldsymbol{\Sigma}}_n &= [\nabla \boldsymbol{\tau}(\hat{\boldsymbol{\theta}}_n)]' \hat{\mathbf{V}}_n [\nabla \boldsymbol{\tau}(\hat{\boldsymbol{\theta}}_n)]. \end{aligned}$$

  Proof: The 1st order Taylor expansion of $\boldsymbol{\tau}(\hat{\boldsymbol{\theta}}_n)$ about $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ gives

$$\boldsymbol{\tau}(\hat{\boldsymbol{\theta}}_n) - \boldsymbol{\tau}(\boldsymbol{\theta}_0) \approx [\nabla \boldsymbol{\tau}(\boldsymbol{\theta}_0)]'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0).$$

  Since $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \approx \mathcal{N}(\mathbf{0}, \hat{\mathbf{V}}_n)$, by linearity of MVN we have

$$\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0 \approx \mathcal{N}(\mathbf{0}, [\nabla \boldsymbol{\tau}(\boldsymbol{\theta}_0)]' \hat{\mathbf{V}}_n [\nabla \boldsymbol{\tau}(\boldsymbol{\theta}_0)]).$$

  Result follows since $\nabla \boldsymbol{\tau}(\hat{\boldsymbol{\theta}}_n) \to \nabla \boldsymbol{\tau}(\hat{\boldsymbol{\theta}}_0)$.

Delta-Method

- **Setup:** Suppose that as $n \to \infty$ we have $\hat{\mathbf{V}}_n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \to \mathcal{N}(\mathbf{0}, \mathbf{I})$, where $(\hat{\boldsymbol{\theta}}_n, \hat{\mathbf{V}}_n)$ are the MLE and inverse Fisher Information calculated from a sequence of random variables $\mathbf{Y}_{1:n}$.

- **Theorem:** Suppose that $\boldsymbol{\tau} : \mathbb{R}^p \to \mathbb{R}^q$ is a differentiable function with $q \le p$, and we wish to estimate $\boldsymbol{\tau}_0 = \boldsymbol{\tau}(\boldsymbol{\theta}_0)$. Then as $n \to \infty$ we have

$$\hat{\boldsymbol{\Sigma}}_n^{1/2}(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0) \to \mathcal{N}(\mathbf{0}, \mathbf{I}), \qquad \begin{aligned} \hat{\boldsymbol{\tau}}_n &= \boldsymbol{\tau}(\hat{\boldsymbol{\theta}}_n) \\ \hat{\boldsymbol{\Sigma}}_n &= [\nabla \boldsymbol{\tau}(\hat{\boldsymbol{\theta}}_n)]' \hat{\mathbf{V}}_n [\nabla \boldsymbol{\tau}(\hat{\boldsymbol{\theta}}_n)]. \end{aligned}$$

- **Upshot:** If $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{V}})$ are the MLE and its variance estimator, a confidence interval for a 1D quantity of interest $\tau_0 = \tau(\boldsymbol{\theta}_0)$ can be constructed via

$$\hat{\tau} \pm 1.96 \cdot s_{\hat{\tau}}, \qquad \begin{aligned} \hat{\tau} &= \tau(\hat{\boldsymbol{\theta}}) \\ s_{\hat{\tau}} &= \sqrt{[\nabla \tau(\hat{\boldsymbol{\theta}})]' \hat{\mathbf{V}} [\nabla \tau(\hat{\boldsymbol{\theta}})]}. \end{aligned}$$

  Can use this to calculate CI for 1-day $\mathrm{VaR}_\tau = \tau(\boldsymbol{\theta}_0) = \mathrm{VaR}_\tau(\boldsymbol{\theta}_0 \mid \boldsymbol{\varepsilon})$.

Value-at-Risk

- **GARCH(1,1) Model:** $$\varepsilon_t = \sigma_t z_t, \qquad z_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$$ $$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$
- **Value-at-Risk:**
$$\mathrm{VaR}_\tau = q_\tau\left(\frac{S_{N+p} - S_N}{S_N} \,|\, \mathbf{S}, \boldsymbol{\theta}\right) \quad \iff \quad \Pr\left(\frac{S_{N+p} - S_N}{S_N} < \mathrm{VaR}_\tau \,|\, \mathbf{S}, \boldsymbol{\theta}\right) = \tau.$$
No analytic solution for $p > 1$.
- **Point Estimate:** Use Monte Carlo:
  1. For given $\boldsymbol{\theta}$, analytically obtain $\sigma_1^2, \ldots, \sigma_N^2$.
  2. Generate M iid realizations of $R = \log(S_{N+p}/S_N)$ from $p(R \,|\, \varepsilon_N, \sigma_N)$ using GARCH. (Note that $R = \sum_{i=1}^{p} \varepsilon_{N+1}$)
  3. The Monte Carlo approximation is $\mathrm{VaR}_\tau = \exp\{\hat{q}_\tau(R \,|\, \varepsilon_N, \boldsymbol{\theta})\} - 1$, where $\hat{q}_\tau(R \,|\, \varepsilon_N, \boldsymbol{\theta})$ is the τ-level sample quantile of the iid realizations $R^{(1)}, \ldots, R^{(M)}$.
- **Interval Estimate:** Use Delta-Method, with $\widehat{\mathrm{VaR}}_\tau = \exp\{\hat{q}_\tau(R \,|\, \varepsilon_N, \hat{\boldsymbol{\theta}})\} - 1$, but with variance reduction, i.e., same $z_{N+1}^{(m)}, \ldots, z_{N+p}^{(m)}$ for every value of $\boldsymbol{\theta}$.

Value-at-Risk

- **GARCH(1,1) Model:**
$$\varepsilon_t = \sigma_t z_t, \qquad z_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$$
$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

- **Value-at-Risk:**
$$\mathrm{VaR}_\tau = q_\tau\left(\frac{S_{N+p} - S_N}{S_N} \mid \mathbf{S}, \boldsymbol{\theta}\right) \quad \iff \quad \Pr\left(\frac{S_{N+p} - S_N}{S_N} < \mathrm{VaR}_\tau \mid \mathbf{S}, \boldsymbol{\theta}\right) = \tau.$$

- **Point/Interval Estimate:** Monte Carlo + Delta Method

- **Model Misspecification:** Suppose we have GARCH(1,1), but with $z_t \overset{\text{iid}}{\sim} F(z)$ with $F \neq \mathcal{N}(0,1)$?

- **Residual Bootstrap:**
  - GARCH model: $\varepsilon_t = \sigma_t z_t, \qquad \sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$
  - Use $\hat{\boldsymbol{\theta}}$ to calculate $\hat{\boldsymbol{\sigma}} = (\hat{\sigma}_1, \ldots, \hat{\sigma}_N)$ and *residuals* $\hat{\mathbf{z}} = (\hat{z}_1, \ldots, \hat{z}_N) = \boldsymbol{\varepsilon}/\hat{\sigma}$.
  - Obtain Bootstrap residuals $\tilde{\mathbf{z}}$ by sampling with replacement from $\hat{\mathbf{z}}$
  - Bootstrap log-returns: $\tilde{\varepsilon}_t = \tilde{\sigma}_t \tilde{z}_t, \qquad \sigma_t^2 = \hat{\omega} + \hat{\alpha} + \tilde{\varepsilon}_{t-1}^2 + \hat{\beta}\tilde{\sigma}_{t-1}^2$