Environmental contours – part 3

Arne Bang Huseby and Kristina Rognlien Dahl

University of Oslo, Norway

STK 4400

The Rosenblatt transformation

The Rosenblatt transformation

As before we consider a vector (T, H) of environmental variables.

The *Rosenblatt transformation*, denoted Ψ, depends on the joint distribution of (T, H), and is such that if:

$$(X, Y) = \Psi(T, H),$$

then X and Y are independent standard normally distributed.

- The *normal space* = The space containing (X, Y).
- The *environmental space* = The space containing (T, H).

The Rosenblatt transformation

The transformation of (T, H) into (X, Y) is done in two steps:

Step 1. Transform (T, H) into (U, V) such that U and V are independent and uniformly distributed on $[0, 1]$.

Step 2. Transform (U, V) into (X, Y)

We let Φ denote the cumulative distribution function of the standard normal distribution. Thus, if X is standard normally distributed, we have:

$$P(X \leq x) = \Phi(x).$$

The cumulative distribution function of H is denoted F_H, while the conditional distribution function of T given H is denoted $F_{T|H}$.

The Rosenblatt transformation (cont.)

Step 1. In this step we let:

$$U = F_H(H),$$

$$V = F_{T|H}(T|H).$$

This implies that for all $u, v \in [0, 1]$ we have:

$$\begin{aligned} P(U \leq u) &= P(F_H(H) \leq u) = P(H \leq F_H^{-1}(u)) \\ &= F_H(F_H^{-1}(u)) = u, \end{aligned}$$

$$\begin{aligned} P(V \leq v) &= P(F_{T|H}(T|H) \leq v) = P(T \leq F_{T|H}^{-1}(v|H)) \\ &= F_{T|H}(F_{T|H}^{-1}(v|H)|H) = v. \end{aligned}$$

Hence, U and V are independent and uniformly distributed on $[0, 1]$.

The Rosenblatt transformation (cont.)

Step 2. In this step we let:

$$X = \Phi^{-1}(U),$$

$$Y = \Phi^{-1}(V).$$

This implies that for all x, y we have:

$$\begin{aligned} P(X \leq x) &= P(\Phi^{-1}(U) \leq x) \\ &= P(U \leq \Phi(x)) = \Phi(x), \end{aligned}$$

$$\begin{aligned} P(Y \leq y) &= P(\Phi^{-1}(V) \leq y) \\ &= P(V \leq \Phi(y)) = \Phi(y). \end{aligned}$$

The inverse Rosenblatt transformation

The *inverse Rosenblatt transformation*, denoted Ψ^{-1} is such that if X and Y are independent standard normally distributed, then

$$(T, H) = \Psi^{-1}(X, Y),$$

then (T, H) have the correct environmental distribution.

The transformation of (X, Y) into (T, H) is done in two steps:

Step 1. Transform (X, Y) into (U, V) such that U and V are independent and uniformly distributed on $[0, 1]$.

Step 2. Transform (U, V) into (T, H)

The inverse Rosenblatt transformation (cont.)

Step 1. In this step we let:

$$U = \Phi(X),$$

$$V = \Phi(Y).$$

This implies that for all $u, v \in [0, 1]$ we have:

$$\begin{aligned} P(U \leq u) &= P(\Phi(X) \leq u) = P(X \leq \Phi^{-1}(u)) \\ &= \Phi(\Phi^{-1}(u)) = u, \end{aligned}$$

$$\begin{aligned} P(V \leq v) &= P(\Phi(Y) \leq v) = P(Y \leq \Phi^{-1}(v)) \\ &= \Phi(\Phi^{-1}(v)) = v. \end{aligned}$$

Hence, U and V are independent and uniformly distributed on $[0, 1]$.

The inverse Rosenblatt transformation (cont.)

Step 2. In this step we let:

$$H = F_H^{-1}(U),$$

$$T = F_{T|H}^{-1}(V|H).$$

This implies that for all h, t we have:

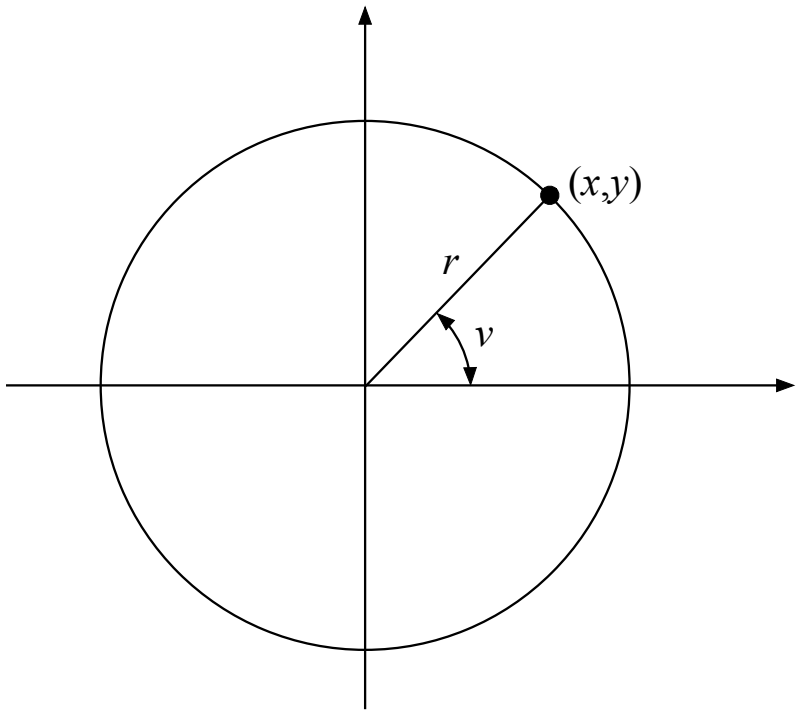
$$\begin{aligned} P(H \leq h) &= P(F_H^{-1}(U) \leq h) \\ &= P(U \leq F_H(h)) = F_H(h), \end{aligned}$$

$$\begin{aligned} P(T \leq t|H = h) &= P(F_{T|H}^{-1}(V|h) \leq t|H = h) \\ &= P(V \leq F_{T|H}(t|h)|H = h) = F_{T|H}(t|h). \end{aligned}$$

Estimating $C(\theta)$ revisited

Improved estimation method

Polar coordinates



If $(x, y) \in \mathbb{R}^2$, then this point can be represented in polar coordinates (r, v), where $r \geq 0$ and $v \in (-\pi, \pi]$ are given by:

$$r = \sqrt{x^2 + y^2}$$

$$v = \text{The angle between the } x\text{-axis and } (x, y)$$

Polar coordinates (cont.)

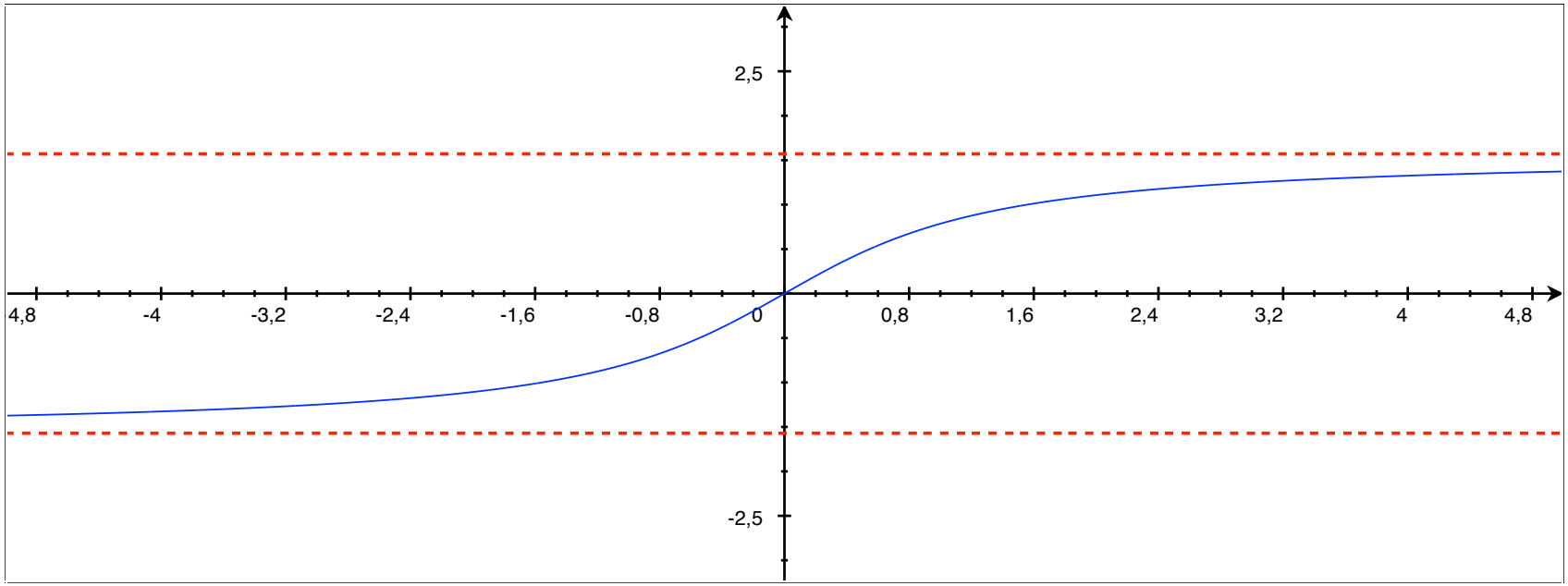
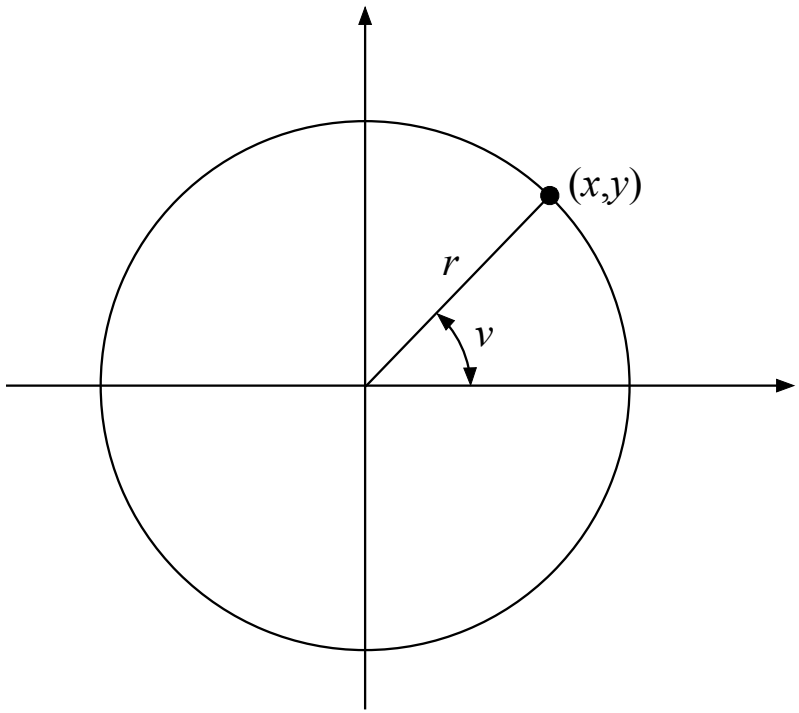


Figure: The $\arctan(z)$-function

We observe that we have $-\frac{\pi}{2} < \arctan(z) < \frac{\pi}{2}$. Thus, $\arctan(z)$ represents an angle between the x-axis and a point (x, y), where $z = y/x$, and where $x > 0$.

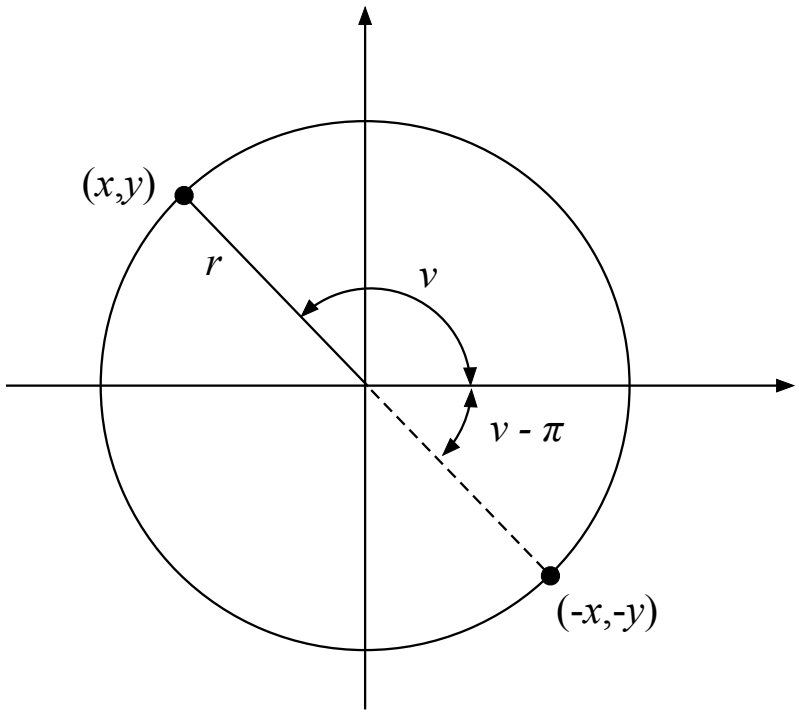
Polar coordinates - Quadrant 1



When $(x, y) \in Q1$, i.e., $x > 0$ and $y > 0$, we have:

$$\arctan(\frac{y}{x}) = v$$

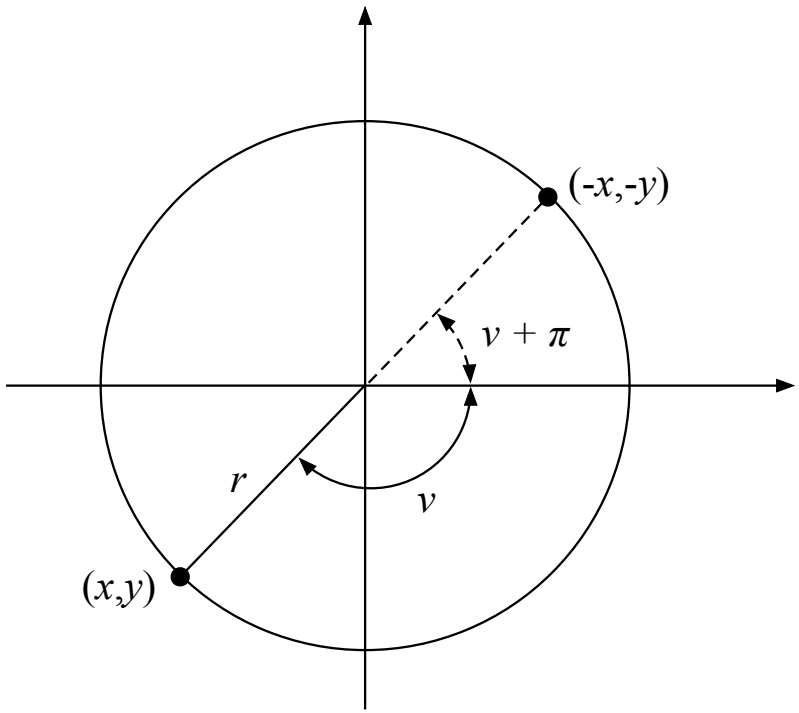
Polar coordinates - Quadrant 2



When $(x, y) \in Q2$, i.e., $x < 0$ and $y > 0$, we have:

$$\arctan(\frac{y}{x}) = v - \pi \quad \Rightarrow \quad v = \arctan(\frac{y}{x}) + \pi$$

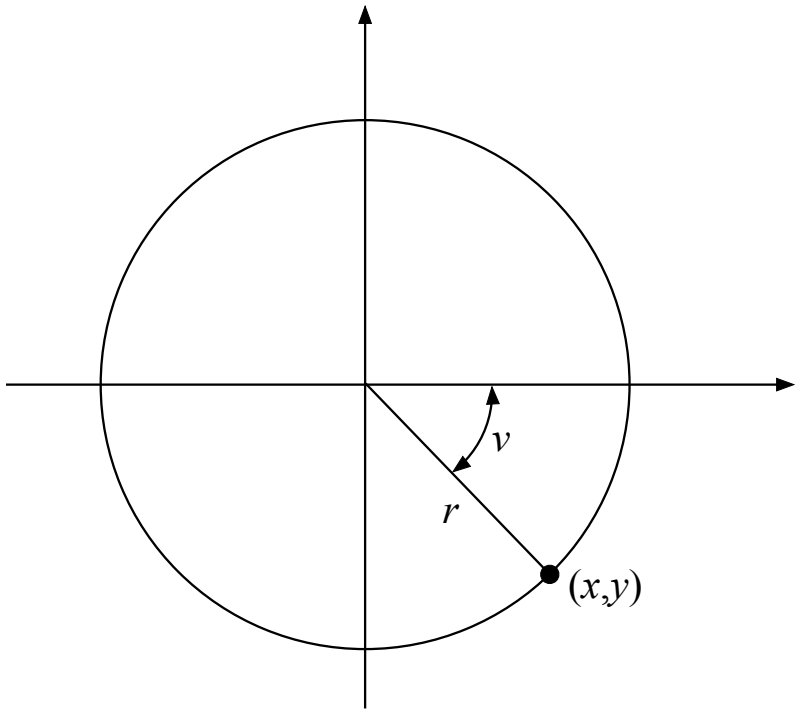
Polar coordinates - Quadrant 3



When $(x, y) \in Q3$, i.e., $x < 0$ and $y < 0$, we have:

$$\arctan(\frac{y}{x}) = v + \pi \quad \Rightarrow \quad v = \arctan(\frac{y}{x}) - \pi$$

Polar coordinates - Quadrant 4



When $(x, y) \in Q4$, i.e., $x > 0$ and $y < 0$, we have:

$$\arctan(\frac{y}{x}) = v$$

Polar coordinates (cont.)

The function $\text{atan2}(x, y)$ is defined as follows:

$$\text{atan2}(x,y) = \begin{cases} \arctan(\frac{y}{x}) & \text{if } x > 0 \text{ and } y \geq 0, \\ \arctan(\frac{y}{x}) + \pi & \text{if } x < 0 \text{ and } y \geq 0, \\ \arctan(\frac{y}{x}) - \pi & \text{if } x < 0 \text{ and } y < 0, \\ \arctan(\frac{y}{x}) & \text{if } x > 0 \text{ and } y < 0, \\ \frac{\pi}{2} & \text{if } x = 0 \text{ and } y > 0, \\ -\frac{\pi}{2} & \text{if } x = 0 \text{ and } y < 0, \\ \text{undefined} & \text{if } x = 0 \text{ and } y = 0. \end{cases}$$

Estimating $C(\theta)$ revisited

Assume that X and Y are independent and normally distributed with mean 0 and standard deviation 1. We then let:

$$R = \sqrt{X^2 + Y^2},$$
$$V = \text{atan2}(X, Y),$$

Thus that R and V are the polar coordinates of (X, Y).

It can be shown that R and V are independent, and:

$$Z = R^2 = X^2 + Y^2 \sim \chi^2_2\text{-distributed}$$

$$V \sim R[-\pi, \pi]\text{-distributed}$$

Estimating $C(\theta)$ revisited (cont.)

The density of Z is given by:

$$f_Z(z) = \frac{1}{2} e^{-z/2}, \quad \text{for } z > 0,$$

which is an exponential distribution with rate $\lambda = 1/2$.

This implies that:

$$P(Z > z) = e^{-z/2}$$

Hence, the probability that (X, Y) is located outside a circle with centrum in origin and with a radius r is equal to $e^{-r^2/2}$.

Estimating $C(\theta)$ revisited (cont.)

To simulate from the distribution of (X, Y) we start by generating U and V, where $U \sim R[0, 1]$ and $V \sim R[-\pi, \pi]$.

We then let $Z = -2\ln(U)$. Now, it is easy to show that Z gets the density f_Z. We also calculate $R = \sqrt{Z}$. Since R and V are the polar coordinates to (X, Y), we find that:

$$X = R\cos(V) = \sqrt{Z} \cdot \cos(V),$$
$$Y = R\sin(V) = \sqrt{Z} \cdot \sin(V).$$

We then let $(T, H) = \Psi^{-1}(X, Y)$, where Ψ^{-1} is the inverse Rosenblatt transformation for the joint distributions of T and H. This way (T, H) gets the correct joint distribution.

Estimating $C(\theta)$ revisited (cont.)

Let $\theta \in [0, 2\pi)$, and let $S(\theta) = T\cos(\theta) + H\sin(\theta)$.

For a given exceedance probability p_e we wish to estimate $C(\theta)$ such that $P(S(\theta) > C(\theta)) = p_e$.

By simulating (T, H) n times, each time calculating the resulting value of $S(\theta)$, we can estimate $C(\theta)$ by the order observator $S_{(k)}(\theta)$, where k is such that:

$$1 - \frac{k}{n} = \frac{n-k}{n} \approx p_e.$$

If p_e is very small, i.e., 0.1%, a large number of simulations are needed in order to obtain stable estimates.

Most of the simulations yield results close to the central area of the joint distribution. Very few of the simulated values provide information about the contour area.

Alternative Monte Carlo sampling scheme

IDEA: Avoid sampling points from the central area of the joint distribution, and just sample points close to the contour.

Thus, we would like to simulate (X, Y) from the conditional distribution for (X, Y) given that this vector falls outside a circle with radius, say r_0.

This is done by sampling (X, Y) from the conditional distribution given that $R = \sqrt{X^2 + Y^2} > r_0$.

Or equivalently, by sampling (X, Y) from the conditional distribution for (X, Y) given that $Z = X^2 + Y^2 > z_0 = r_0^2$.

Alternative Monte Carlo sampling scheme (cont.)

The conditional distribution for Z given that $Z > z_0$ is given by:

$$P(Z > z | Z > z_0) = \frac{P(Z > z \bigcap Z > z_0)}{P(Z > z_0)} = \frac{P(Z > z)}{P(Z > z_0)} = e^{-(z - z_0)/2}.$$

Hence, given that $Z > z_0$, $(Z - z_0)$ is exponentially distributed with $\lambda = 1/2$.

Thus, we can simulate from the conditional distribution for Z given $Z > z_0$ by generating $U \sim R[0, 1]$ and let:

$$Z = z_0 - 2\ln(U) = r_0^2 - 2\ln(U).$$

As before, the angle V is generated from the $R[-\pi, \pi]$-distribution.

Finally, we let $X = \sqrt{Z}\cos(V)$ and $Y = \sqrt{Z}\sin(V)$, and $(T, H) = \Psi^{-1}(X, Y)$, where Ψ^{-1} is the inverse Rosenblatt transformation.

Alternative Monte Carlo sampling scheme (cont.)

NOTE: The simulations are focused in the area of interest on the outer edge of the outcome space where we expect that the contour is.

However, we need to correct for this by estimating the percentile function $C(\theta)$ using an *adjusted* exceedance probability which takes into acount that we are not simulating from the true joint distributions of T and H.

We let $p'_e = P(S(\theta) > C(\theta)|R > r_0)$ be this adjusted exceedance probability, and assume that r_0 is chosen such that the event $\{S(\theta) > C(\theta)\}$ is contained in the event $\{R > r_0\}$.

We can achieve this by ensuring that r_0 is *not too large*.

Alternative Monte Carlo sampling scheme (cont.)

Assuming that the event $\{S(\theta) > C(\theta)\}$ is contained in the event $\{R > r_0\}$, we have:

$$p'_e = P(S(\theta) > C(\theta) | R > r_0) = \frac{P(S(\theta) > C(\theta) \bigcap R > r_0)}{P(R > r_0)}$$

$$= \frac{P(S(\theta) > C(\theta))}{P(R > r_0)}$$

$$= \frac{p_e}{e^{-r_0^2/2}} = e^{r_0^2/2} \cdot p_e,$$

where we have used that:

$$P(R > r_0) = P(Z > r_0^2) = e^{-r_0^2/2}.$$

Alternative Monte Carlo sampling scheme (cont.)

We can then simulate n times from this conditional distribution and estimate $C(\theta)$ by the order observation $S_{(k')}(\theta)$, but where k' is determined so that:

$$1 - \frac{k'}{n} = \frac{n - k'}{n} \approx p'_e = e^{r_0^2/2} \cdot p_e.$$

NOTE: Since $r_0 > 0$, we have $e^{r_0^2/2} > 1$.

Hence, $p'_e > p_e$ and $k' < k$. This means that a (much) larger fraction of the simulated data is used to estimate $C(\theta)$.

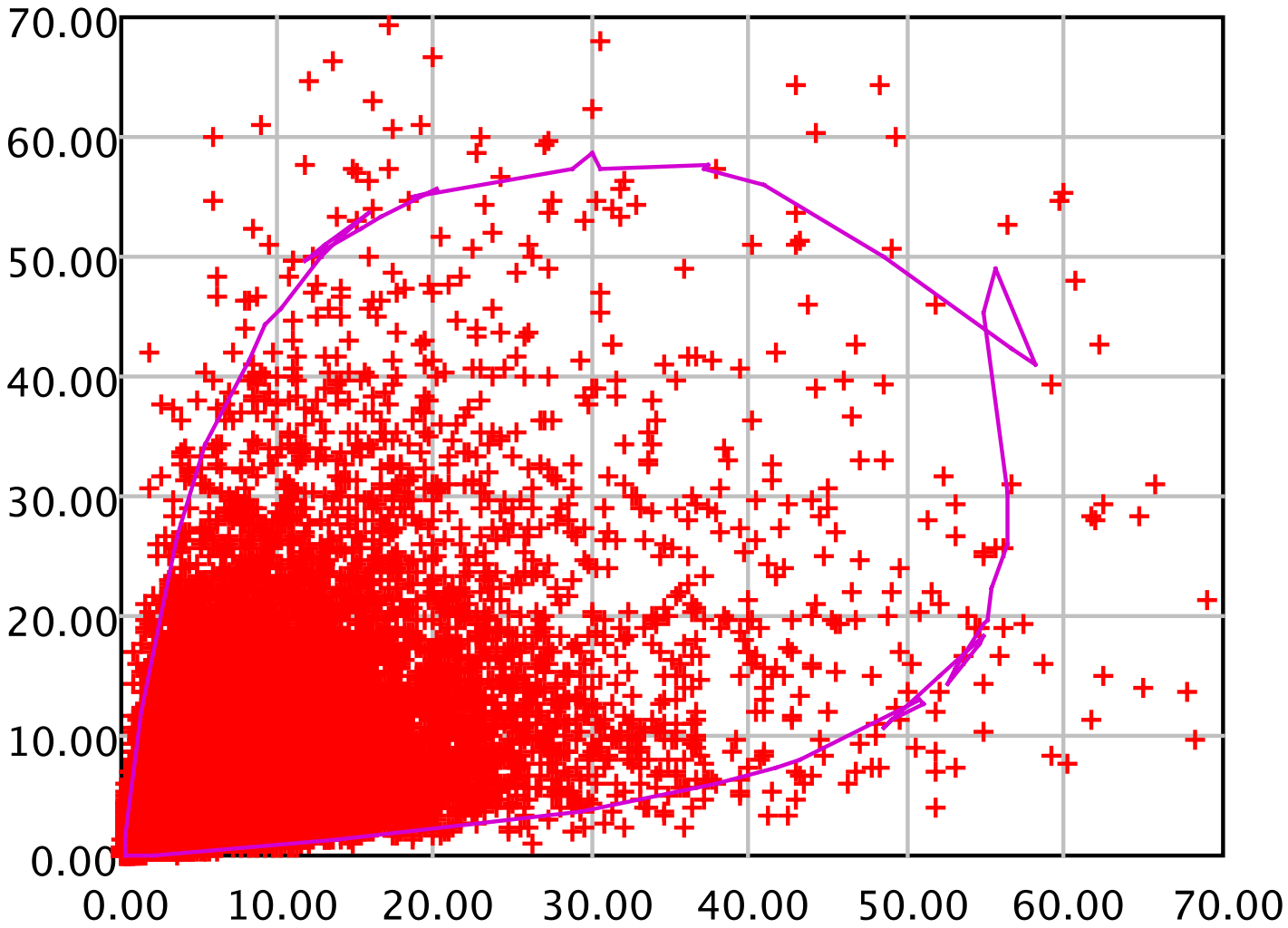
Alternative Monte Carlo sampling scheme (cont.)

Ideally, we would like r_0 to be as large as possible to maximize the effect of the importance sampling. At the same time we must ensure that the event $\{S(\theta) > C(\theta)\}$ is contained in the event $\{R > r_0\}$.

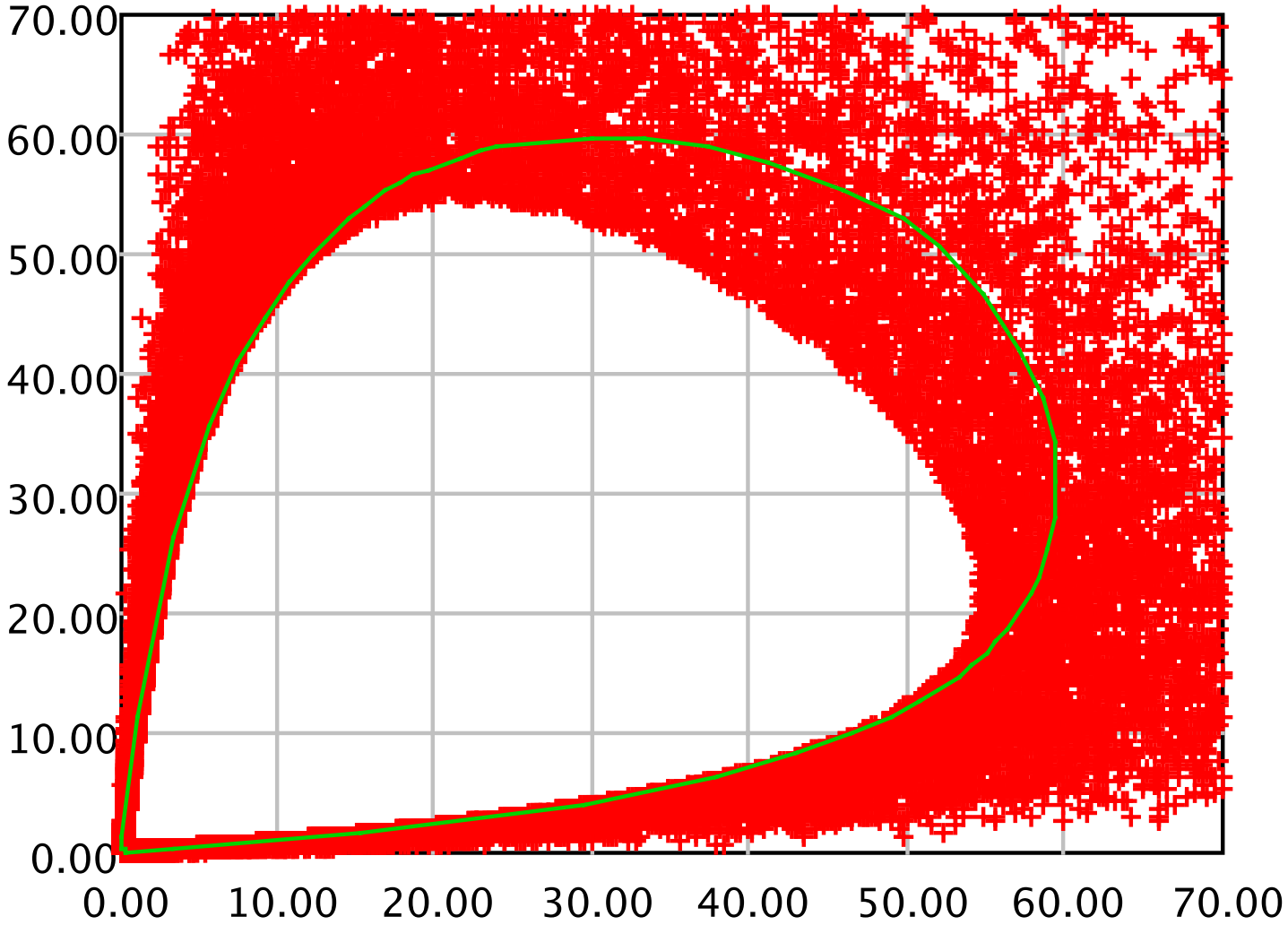
We let $\mathcal{O}$ denote a circle centered in the origin with radius r_0. Then r_0 must be chosen so that the transformed set $\Psi^{-1}(\mathcal{O})$ is contained inside the contour we want to estimate.

Experiences have shown that we get a stable estimate by choosing $r_0 = 0.95 \cdot r$, where r is the $(1 - p_e)$-percentile in the standard normal distribution, and p_e is the target exceedance probability..

Estimating a contour *without* importance sampling



Estimating a contour *with* importance sampling



Transformed contours

Transformed contours

As already explained, the *Rosenblatt transformation* Ψ is such that if:

$$(T', H') = \Psi(T, H),$$

then T' and H' are independent standard normally distributed.

We also recall that:

- The *normal space* = The space containing (T', H').
- The *environmental space* = The space containing (T, H).

For a given set $\mathcal{B}'$ in the *normal space*, we let $\mathcal{E}'$ be the family of all convex sets $\mathcal{F}'$ in the *normal space* such that $\mathcal{F}' \cap \mathcal{B}' \subseteq \partial\mathcal{B}'$.

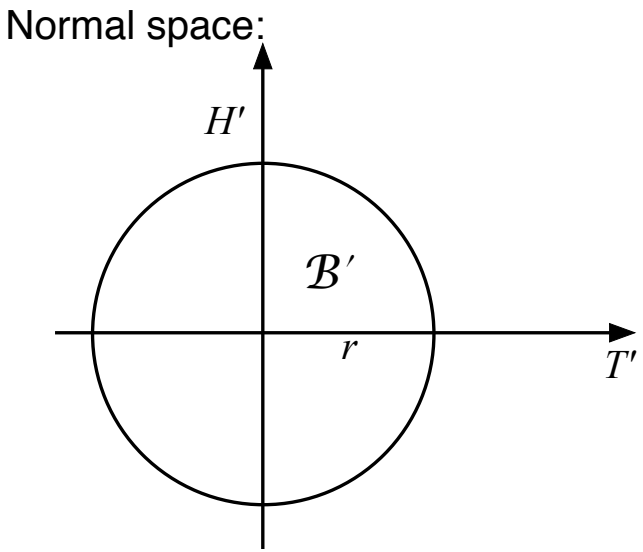
If $\mathcal{B}'$ is convex, it follows by the halfspace failure region proposition that:

$$\mathcal{E}'^{*} = \mathcal{P}^{+}(\mathcal{B}').$$

Transformed contours (cont.)

Let $p_e < 0.5$ be the desired exceedance probability, and let $r > 0$ denote the $(1 - p_e)$-percentile in the standard normal distribution.

A contour $\partial\mathcal{B}'$ for (T', H') is constructed by letting $\mathcal{B}'$ be a circle centered at the origin and with radius r.



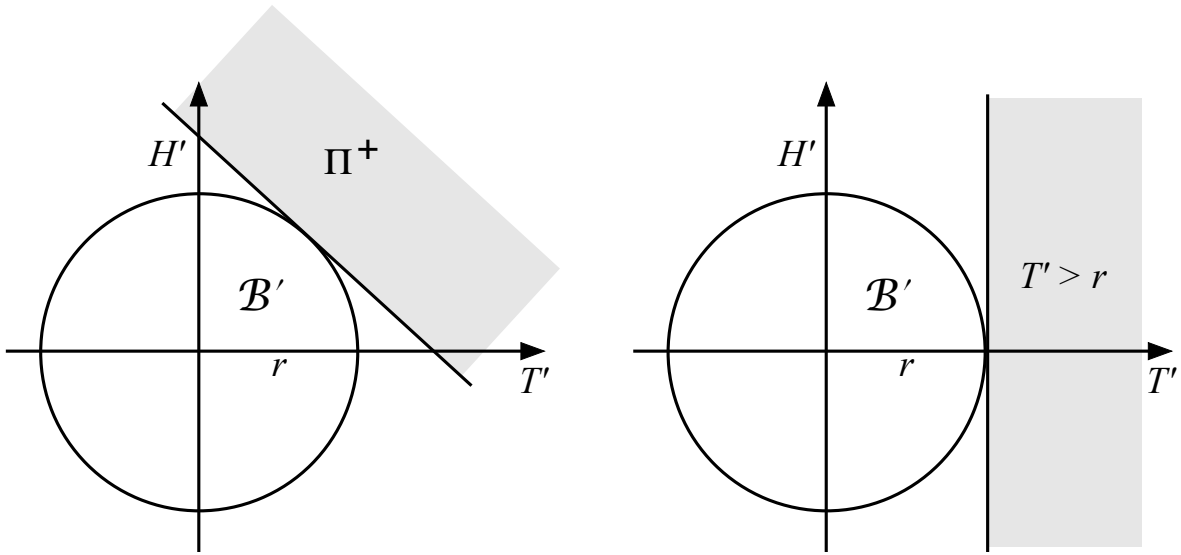
NOTE: Since T' and H' are standard normally distributed, it follows that:

$$P[T' > r] = P[H' > r] = p_e.$$

Transformed contours (cont.)

If $\Pi^+ \in \mathcal{P}^+(\mathcal{B}')$, it follows by the rotational symmetry property of normal distribution that:

$$P[(T', H') \in \Pi^+] = P[T' > r] = p_e.$$



Since this is true for all $\Pi^+ \in \mathcal{P}^+(\mathcal{B}')$, we then get:

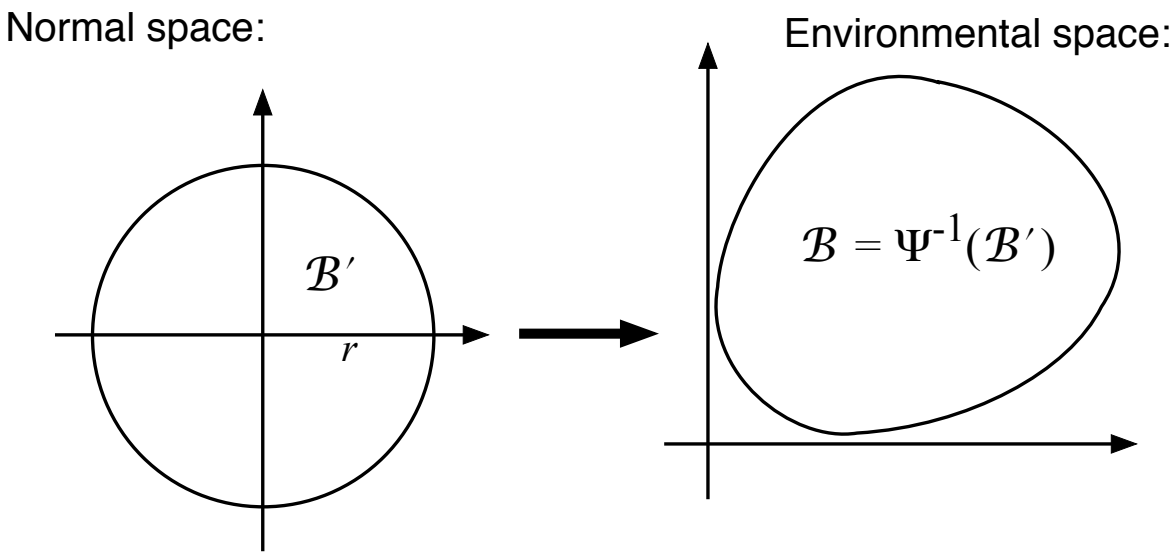
$$P_e(\mathcal{B}', \mathcal{E}') = \sup\{P[(T', H') \in \Pi^+] : \Pi^+ \in \mathcal{P}^+(\mathcal{B}')\} = p_e.$$

Transformed contours (cont.)

A contour set $\mathcal{B}$ is then obtained by transforming the set $\mathcal{B}'$ from normal space back to the environmental space using the *inverse Rosenblatt transformation*.

That is, we let $\mathcal{B}$ be given by:

$$\mathcal{B} = \Psi^{-1}(\mathcal{B}') = \{(t, h) = \Psi^{-1}(x, y) : (x, y) \in \mathcal{B}'\}$$



Exceedance probability of transformed contours

Estimating the exceedance probability of transformed contours

Convex failure regions in the normal space

In the environmental space we may compute the exceedance probability of the contour with respect to the family of *transformed normal space failure regions*. That is, we let:

$$\mathcal{E} = \{\mathcal{F} = \Psi^{-1}(\mathcal{F}') : \mathcal{F}' \in \mathcal{E}'\},$$

For this family of failure regions we get:

$$\begin{aligned} P_e(\mathcal{B}, \mathcal{E}) &= \sup\{P[(T, H) \in \mathcal{F}] : \mathcal{F} \in \mathcal{E}\} \\ &= \sup\{P[(T, H) \in \Psi^{-1}(\mathcal{F}')] : \mathcal{F}' \in \mathcal{E}'\} \\ &= \sup\{P[(T', H') \in \mathcal{F}'] : \mathcal{F}' \in \mathcal{E}'\} \\ &= \sup\{P[(T', H') \in \mathcal{F}'] : \mathcal{F}' \in \mathcal{E}'^{*}\} \\ &= \sup\{P[(T', H') \in \Pi^{+}] : \Pi^{+} \in \mathcal{P}^{+}(\mathcal{B}')\} = p_e. \end{aligned}$$

Hence, the contour $\partial\mathcal{B}$ indeed has the desired exceedance probability with respect to this particular family of failure regions.

Exceedance probability of transformed contours

PROBLEM 1: We observe that $\mathcal{E}$ consists of transformed normal space failure regions, where the transformation depends on the joint distribution of (T, H).

Environmental conditions may vary a lot from location to location. If the family $\mathcal{E}$ depends on the joint distribution of (T, H), $\mathcal{E}$ also varies from location to location.

The true failure region of a given mechanical construction, however, should be the same irrespective of location.

PROBLEM 2: The transformed normal space failure regions will in general *not* be convex.

In the following we instead assume that $\mathcal{E}$ is the family of convex sets $\mathcal{F}$ such that $\mathcal{F} \cap \mathcal{B} \subseteq \partial\mathcal{B}$.

Convex failure regions in the environmental space

We recall that the *exceedance probability* of $\mathcal{B}$ with respect to $\mathcal{E}$ is defined as:

$$P_e(\mathcal{B}, \mathcal{E}) = \sup\{P[(T, H) \in \mathcal{F}] : \mathcal{F} \in \mathcal{E}\}.$$

Since a transformed set $\mathcal{B}$ may not itself be convex, the family of maximal failure regions, $\mathcal{E}^*$ is in general not equal to $\mathcal{P}^+(\mathcal{B})$. Hence, it turns out to be difficult to go through all sets $\mathcal{F} \in \mathcal{E}^*$ in order to identify the set with the highest probability.

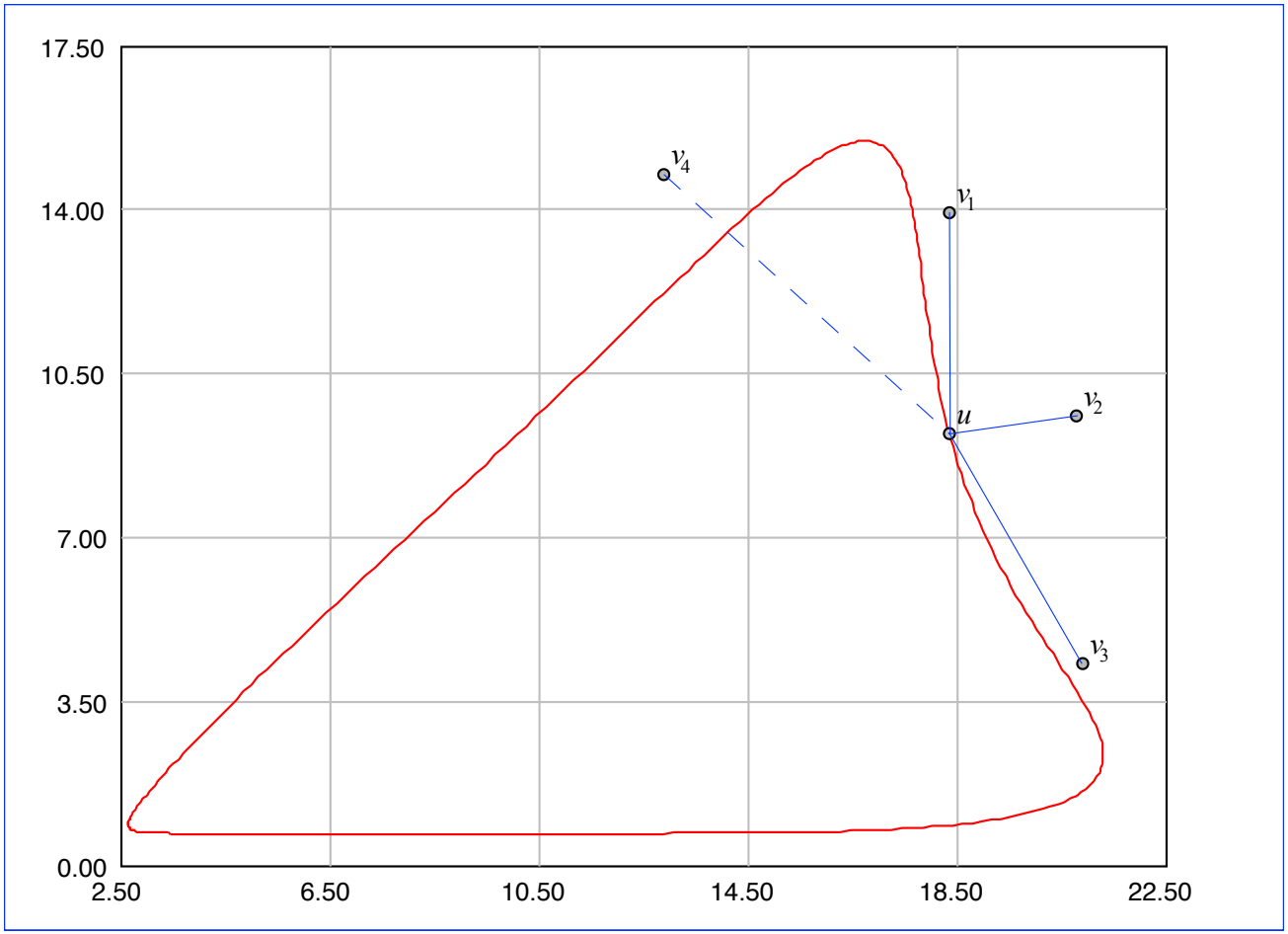
Instead we work with a slightly modified family of failure regions denoted $\tilde{\mathcal{E}}$, defined as follows:

$$\tilde{\mathcal{E}} = \{\tilde{\mathcal{F}}(u) : u \in \partial\mathcal{B}\},$$

where $\tilde{\mathcal{F}}(u)$ is the set of all points $v \notin \mathcal{B}$ that are *visible* from u.

A point $v \notin \mathcal{B}$ is said to be visible from u if the line between u and v does not intersect the interior of $\mathcal{B}$.

Visible points



From the point u, the three points v_1, v_2 and v_3 are visible, while the point v_4 is *not* visible.

Visibility method

Proposition

Assume that $\mathcal{F} \in \mathcal{E}$. Then there exists a set $\tilde{\mathcal{F}} \in \tilde{\mathcal{E}}$ such that $\mathcal{F} \subseteq \tilde{\mathcal{F}}$.

PROOF: Let $\mathcal{F} \in \mathcal{E}$. Then there exists a set $\mathcal{F}^* \in \mathcal{E}^*$ such that $\mathcal{F} \subseteq \mathcal{F}^*$. Then there must exist at least one point $u \in \partial\mathcal{B}$ such that $u \in \mathcal{F}^*$. If this is not the case, this contradicts that $\mathcal{F}^*$ is a maximal failure region.

Now, let $v \in \mathcal{F}^*$ be arbitrary. Since $\mathcal{F}^*$ is convex, the line segment between u and v is contained inside $\mathcal{F}^*$.

Since $\mathcal{F}^* \cap \mathcal{B} \subseteq \partial\mathcal{B}$, the line segment between u and v does not intersect the interior of $\mathcal{B}$.

Hence, v is visible from u, and since v was chosen arbitrarily, this implies that all points in $\mathcal{F}^*$ are visible from u.

Thus, by letting $\tilde{\mathcal{F}} = \tilde{\mathcal{F}}(u) \in \tilde{\mathcal{E}}$, we get that:

$$\mathcal{F} \subseteq \mathcal{F}^* \subseteq \tilde{\mathcal{F}}.$$

Visibility method (cont.)

Corollary

Let $\mathcal{B}$ be an environmental contour set. Then we have:

$$P_e(\mathcal{B}, \mathcal{E}) \leq P_e(\mathcal{B}, \tilde{\mathcal{E}})$$

PROOF: Let $\mathcal{F} \in \mathcal{E}$. Then by the above proposition there exists a set $\tilde{\mathcal{F}} \in \tilde{\mathcal{E}}$ such that $\mathcal{F} \subseteq \tilde{\mathcal{F}}$.

Hence, we have:

$$P[(T, H) \in \mathcal{F}] \leq P[(T, H) \in \tilde{\mathcal{F}}]$$

From this it follows that:

$$\begin{aligned} P_e(\mathcal{B}, \mathcal{E}) &= \sup\{P[(T, H) \in \mathcal{F}] : \mathcal{F} \in \mathcal{E}\} \\ &\leq \sup\{P[(T, H) \in \tilde{\mathcal{F}}] : \tilde{\mathcal{F}} \in \tilde{\mathcal{E}}\} \\ &= P_e(\mathcal{B}, \tilde{\mathcal{E}}). \end{aligned}$$

Visible points - example 1

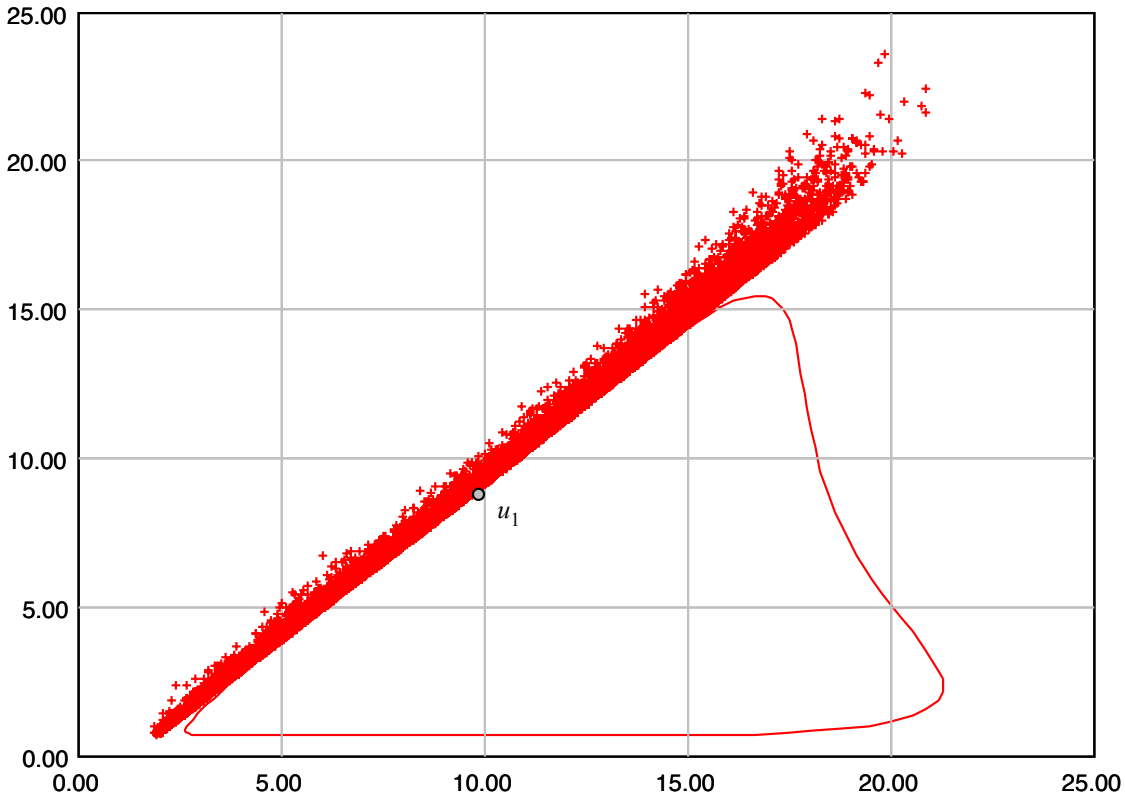


Figure: Estimation of $P((T,H) \in \tilde{\mathcal{F}}(u_1))$ using simulated visible points

Visible points - example 2

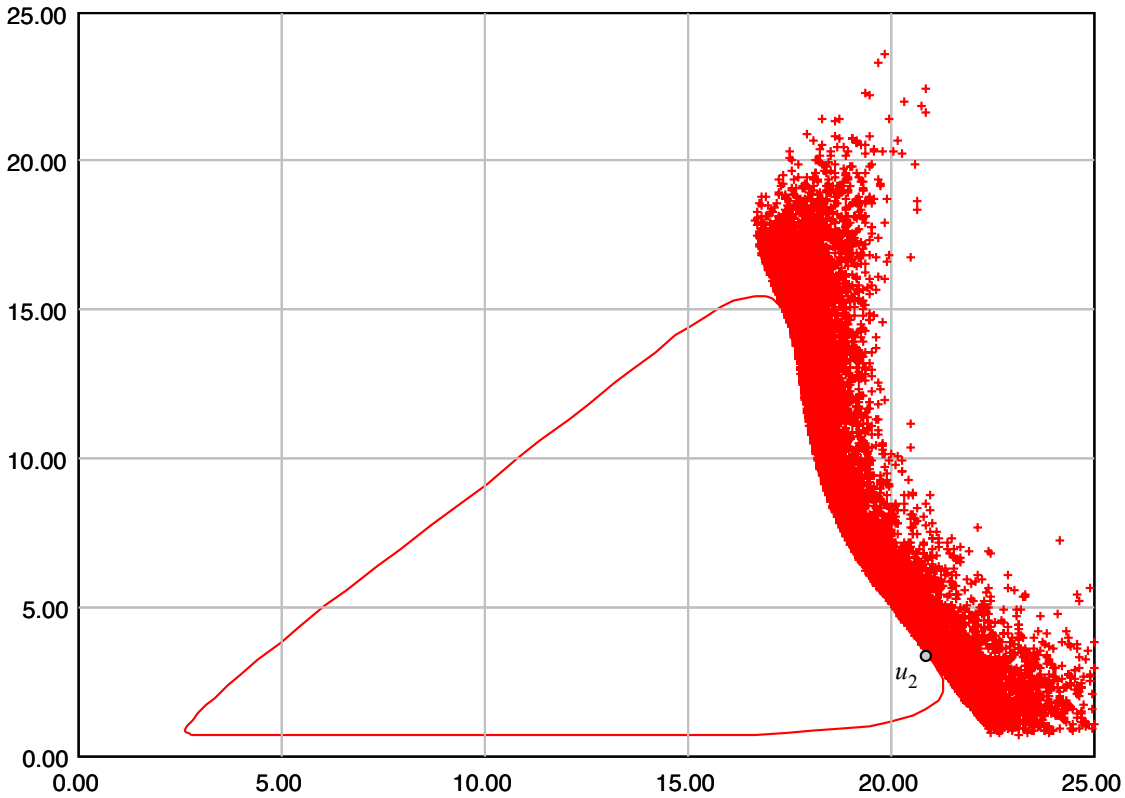


Figure: Estimation of $P((T, H) \in \tilde{\mathcal{F}}(u_2))$ using simulated visible points

Visibility method (cont.)

It can be shown that if $\mathcal{B}$ is *convex* and $\partial\mathcal{B}$ is a differentiable curve, we get that:

$$P_e(\mathcal{B}, \mathcal{E}) = P_e(\mathcal{B}, \tilde{\mathcal{E}}).$$

Thus, in such cases $P_e(\mathcal{B}, \tilde{\mathcal{E}})$ is exact. Furthermore, in general the sets in $\tilde{\mathcal{E}}$ are *almost convex*. Thus, $P_e(\mathcal{B}, \tilde{\mathcal{E}})$ is typically a very good upper bound on the true exceedance probability.

An efficient algorithm for estimating $P_e(\mathcal{B}, \tilde{\mathcal{E}})$ is given in Huseby et al (2019).

Numerical example (cont.)

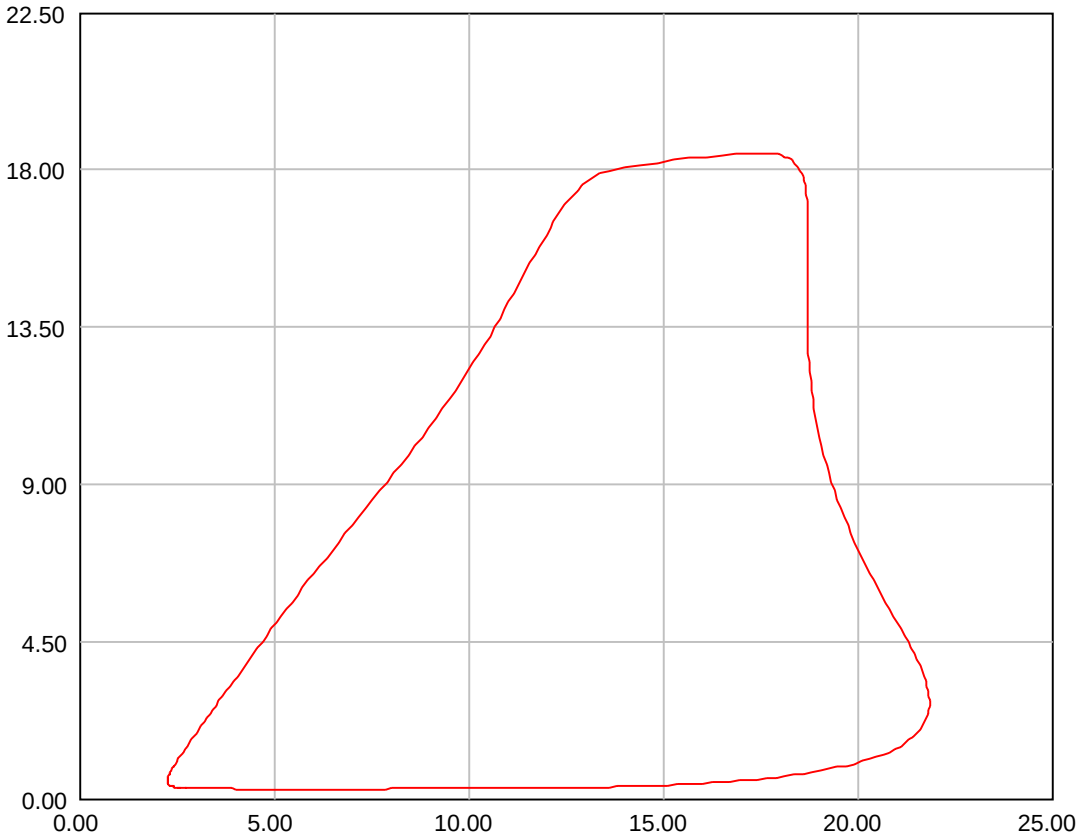


Figure: Environmental contour for the mixed bivariate distribution.

Numerical example (cont.)

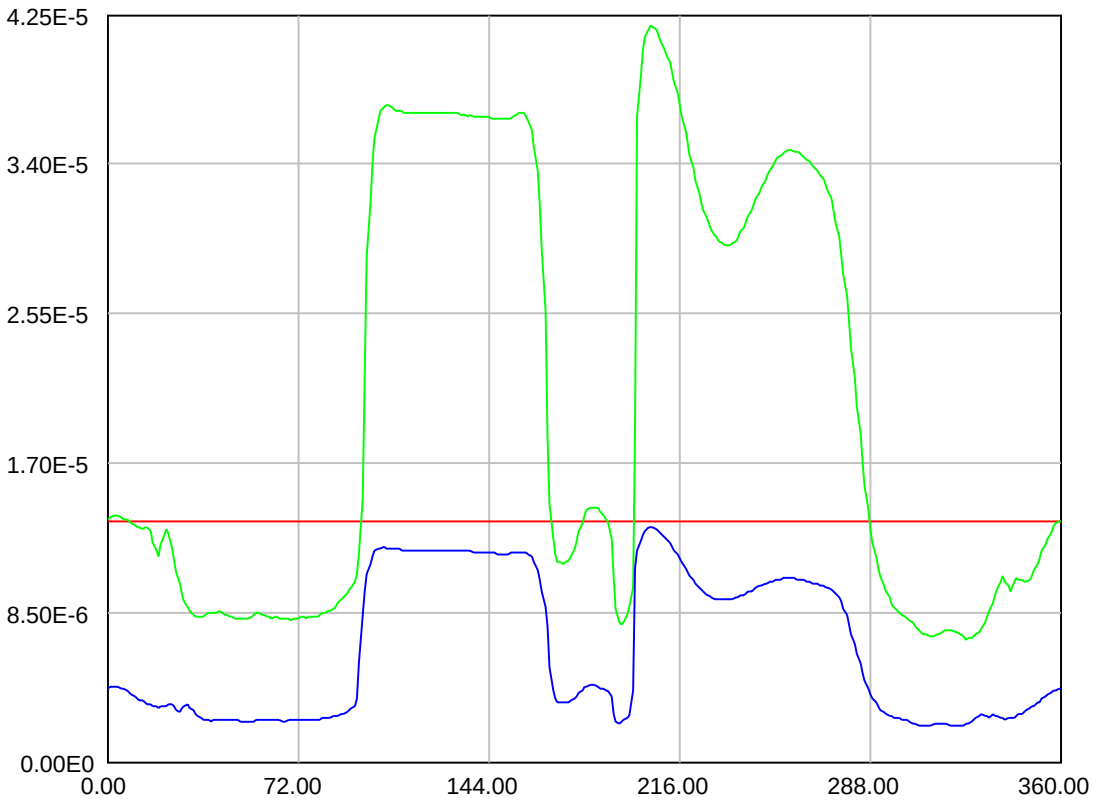


Figure: Estimated probabilities for the sets $\tilde{\mathcal{F}}(u_i)$, $i = 1, \ldots, 360$ for the original environmental contour (green), the desired exceedance probability (red) and an adjusted environmental contour (blue)

Numerical example (cont.)

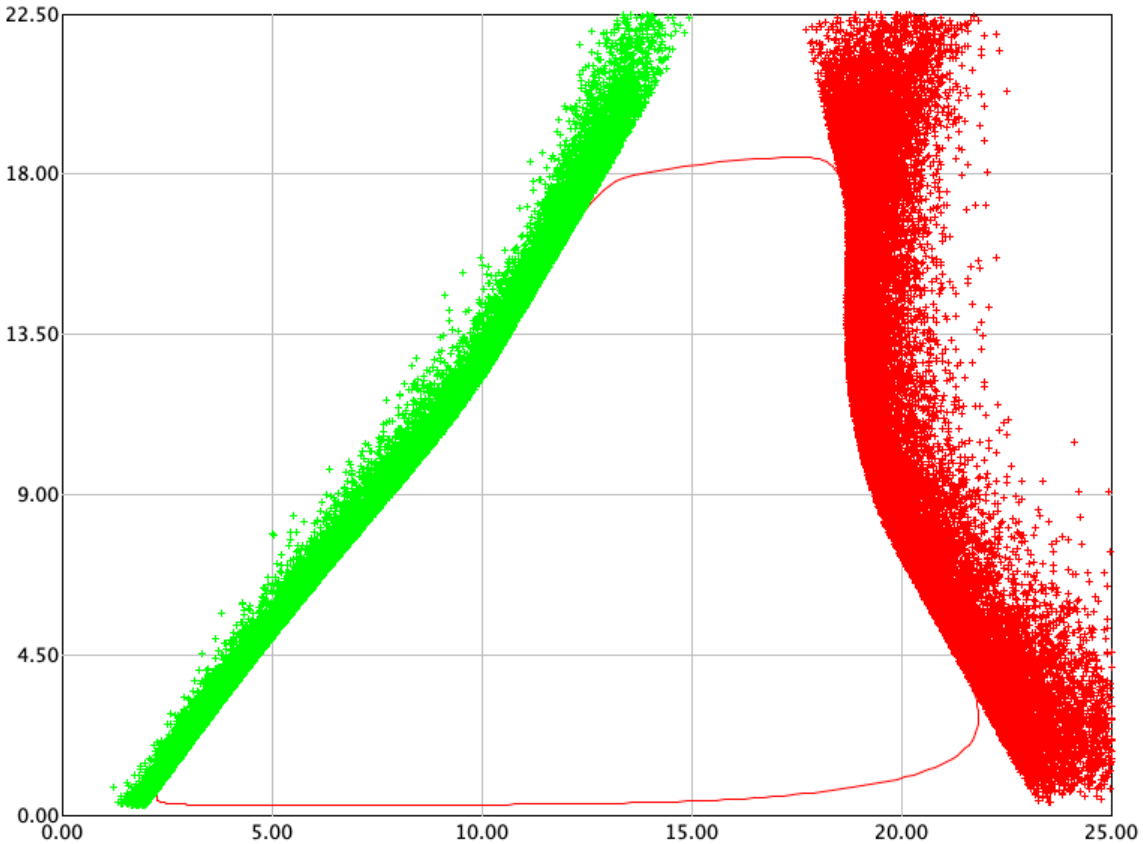


Figure: Environmental contour along with simulated outcomes in the sets $\tilde{\mathcal{F}}(u_{105})$ (red scatter) and $\tilde{\mathcal{F}}(u_{205})$ (green scatter).

Numerical example (cont.)

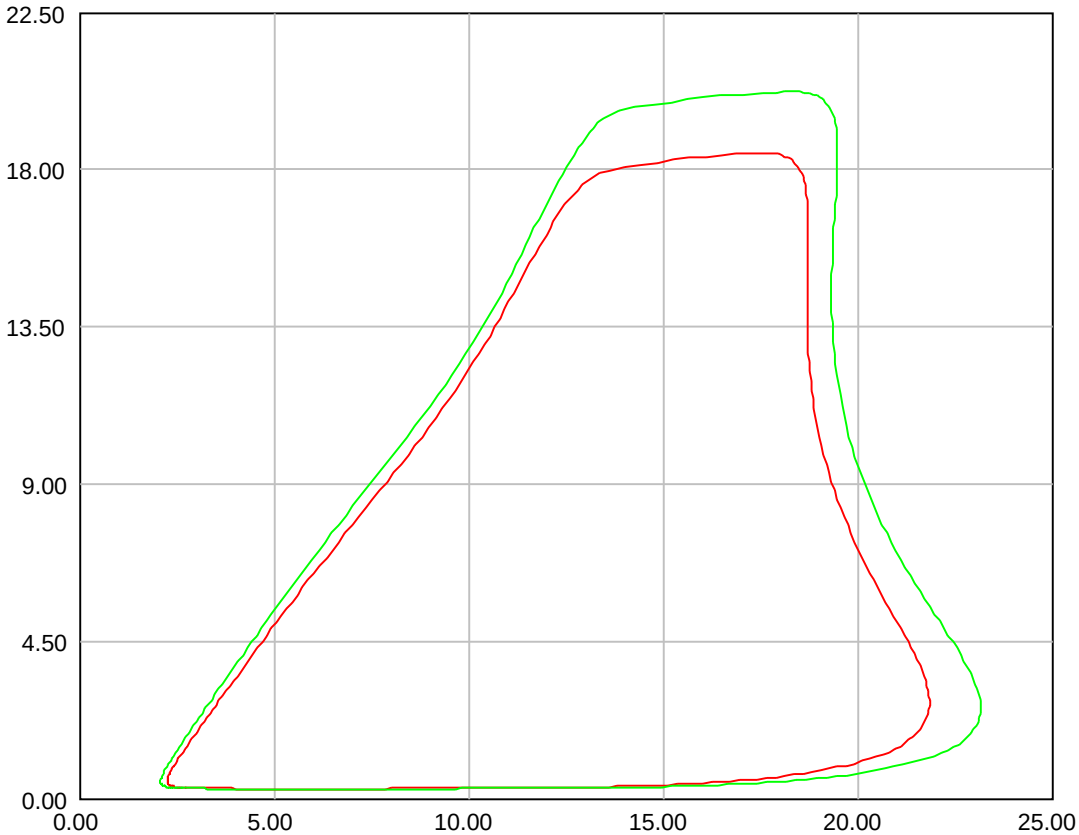


Figure: Original (red curve) and adjusted (green curve) environmental contours for the mixed bivariate distribution.