Poisson approximation

Durrett gives a couple of examples in which a Poisson approximation may be obtained even when a strict independence assumption does not hold. Instead of these *ad hoc* examples, I would like to tell you about a theorem proved in the late 1980's that allows you to quantify the distance from independence and often to derive a Poisson limit for "weakly dependent" random variables. The abstract method is due to L. Chen and C. Stein, but the concrete version is due to Arratia, Goldstein and Gordon. I have posted two versions of their seminal paper on the web, one the original research publication[1] and one from a review the following year[2].

Let I be a finite or countably infinite index set. Let $\{X_\alpha : \alpha \in I\}$ be random variables, each taking value zero or one. For each $\alpha \in I$ we choose a set B_α, which should be thought of as the set of indices β such that X_α and X_β are highly dependent. Note that this is not a precise heuristic, and that the theorem below will be true no matter what you pick for B_α – it will be informative only if $\{B_\alpha\}$ is chosen with some care. Let p_α denote $\mathbb{E} X_\alpha$ and $p_{\alpha,\beta} := \mathbb{E} X_\alpha X_\beta$. Now define the following quantities.

$$b_1 := \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} p_\alpha p_\beta \, , \tag{0.1}$$

$$b_2 := \sum_{\alpha \in I} \sum_{\alpha \neq \beta \in B_\alpha} p_{\alpha,\beta} \tag{0.2}$$

$$b_3 = \sum_{\alpha \in I} \mathbb{E} \left| \mathbb{E} \left(X_\alpha \mid X_\beta : \beta \notin B_\alpha \right) - p_\alpha \right| \, . \tag{0.3}$$

Loosely speaking, b_1 measures the total size of the dependence neighborhoods, b_2 measures how many dependent pairs are likely to arise, and b_3 measures how honest we were in creating the dependency neighborhood (how far X_α is from being independent of $\{X_\beta : \beta \notin B_\alpha\}$).

[1] Arratia, R., Goldstein, L. and Gordon, L. (1989) "Two moments suffice for Poisson Approximation", *Ann. Prob.* **17**:9–25.

[2] Arratia, R., Goldstein, L. and Gordon, L. (1990) "Poisson Approximation and the Chen-Stein method", *Stat. Sci.* **5**:403–424.

Theorem 1 (Arratia-Goldstein-Gordon (1989)) *Let* $W = \sum_{\alpha \in I} X_\alpha$ *and let* Z *be a Poisson random variable with mean* $\lambda := \mathbb{E}W = \sum_{\alpha \in I} p_\alpha$. *Then the total variation distance between* W *and* Z *is at most* $b_1 + b_2 + b_3$.

Note: consult the articles for a stronger conclusion in the case where λ is large, where the upper bound on the total variation distance is multiplied by $1/\lambda$.

Example: the birthday problem. What is the probability that no two people in a room share the same birthday? Assume N people with birthdays IID among the 365 days of the (non-leap) year (below, it will be clear how to replace 365 by any number). As pointed out in [AGG90], the exact answer may be computed in this case, but the Poisson approximation is much more robust, extending to multiple-person birthday coincidences, birthday clusters in consecutive days, and many other variations.

Let I denote the set of pairs of people, so $|I| = \binom{N}{2}$, and for $\alpha \in I$, let X_α be the indicator function of the event that the two people have the same birthday. Let B_α be the set of pairs β not disjoint from α. This makes $b_3 = 0$: it is easy to verify that knowing all coincidences of birthdays among $N - 2$ people, does not alter the chance that the other two people share a birthday. We know that $p_\alpha = 1/365$ for all α and that for any distinct α and β, $p_{\alpha,\beta} = 1/365^2$. Consequently, $b_1 = |I||B_\alpha|365^{-2} = \binom{N}{2}(2N-1)365^{-2}$.

The mean number of coincidences is $\binom{N}{2}365^{-1}$, so we see that the interesting range for N is order $365^{1/2}$, making the Poisson not near zero or one. Substituting $N = \lambda\sqrt{365}$ gives $b_1 \leq \lambda^3 365^{-1/2}$. Observing in this case that $b_2 \leq b_1$ gives $||W - Z|| \leq 2\lambda^3 365^{-1/2}$. This problem and some of the generalizations mentioned above are worked out in [AGG90].